XXXII. MAGYAR OPERÁCIÓKUTATÁSI KONFERENCIA
CEGLÉD, 2017. JÚNIUS 14-16.

Valószínűség-maximalizálás belső közelítéssel

Fábián Csaba, Csizmás Edit, Drenyovszki Rajmund,
Win van Ackooij, Vajnai Tibor, Kovács Lóránt,
Szántai Tamás

Pallasz
Athéné
Egyetem

Outline

Probabilistic optimization problems

Classic and recent solution approaches

A variant of a classic approach: epi-approximation
- Approximating the epigraph of the probabilistic function

Computational study

Theoretical justification of heuristic adjustments

Working with gradient estimates

Problem formulation

$\boldsymbol{\xi}$ an n-dimensional random vector, with known distribution.

Let $F(.)$ denote the cumulative distribution function.

$\boldsymbol{x} \in \mathbb{R}^m$ decision variables

$T\boldsymbol{x}$ a linear function of the decision variables

Probabilistic function

$$\mathrm{P}(T\boldsymbol{x} \geq \boldsymbol{\xi}) = F(T\boldsymbol{x})$$

Probability maximization

$$\max\ F(T\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{x} \in X.$$

Joint probabilistic constraint

$$\min\ \boldsymbol{c}^T\boldsymbol{x} \quad \text{subject to} \quad \boldsymbol{x} \in X, \qquad F(T\boldsymbol{x}) \geq p.$$

Equivalent formulation for joint probabilistic constraint:

$$\min\ \boldsymbol{c}^T\boldsymbol{x} \quad \text{subject to} \quad \boldsymbol{x} \in X, \qquad T\boldsymbol{x} \in \mathcal{L},$$

where $\mathcal{L} = \{\, \boldsymbol{z} \mid F(\boldsymbol{z}) \geq p \,\}.$

Early application

strategic planning model for the Hungarian energy sector
Prékopa, Ganczer, Deák, Patyi (1980).

Characterization and convexity statements, Prékopa (1970-1973):

Assume $\boldsymbol{\xi}$ has a continuous distribution with a log-concave density function.

$\Rightarrow$ the cumulative distribution function $F(.)$ is log-concave

$\Rightarrow$ the probabilistic function $\boldsymbol{x} \mapsto F(T\boldsymbol{x})$ is log-concave.

Solution approaches

Feasible direction method in the convex level set $\mathcal{L}$.

Prékopa, Ganczer, Deák, Patyi (1980).

Cutting-plane methods approximating the convex level set $\mathcal{L}$.

Prékopa and Szántai (1978), Szántai (1988), Mayer (1998),

Henrion and associates (2000 -).

Solution approaches

Feasible direction method in the convex level set $\mathcal{L}$.

Prékopa, Ganczer, Deák, Patyi (1980).

Cutting-plane methods approximating the convex level set $\mathcal{L}$.

Prékopa and Szántai (1978), Szántai (1988), Mayer (1998),

Henrion and associates (2000 -).

- Efficiency due to reusing former gradient information.
- Difficulty: gradient computation is noisy.
  Practicable implementations require
  sophisticated tolerance handling.

Other solution approaches

Uncertain convex programs,

Campi, Calafiore, Garatti, Caré (2005 -)

Sample Average Approximation, integer programming formulations,

Ahmed, Luedtke, Nemhauser (2007 -)

Cone programming,

Cheng, Gicquel and Lisser (2012).

Classic dual approach

Idea: to build an inner approximation of the level set $\mathcal{L}$.

Classic dual approach

Idea: to build an inner approximation of the level set $\mathcal{L}$.

Definition (Prékopa 1990): z is a p-efficient point iff

$$\begin{cases} z \in \mathcal{L} \quad \text{and} \\ \text{there exists no } z' \in \mathcal{L} \text{ such that } z' \leq z,\, z' \neq z. \end{cases}$$

Classic dual approach

Idea: to build an inner approximation of the level set $\mathcal{L}$.

Definition (Prékopa 1990): $\boldsymbol{z}$ is a p-efficient point iff

$$\begin{cases} \boldsymbol{z} \in \mathcal{L} \quad \text{and} \\ \text{there exists no } \boldsymbol{z}' \in \mathcal{L} \text{ such that } \boldsymbol{z}' \leq \boldsymbol{z}, \boldsymbol{z}' \neq \boldsymbol{z}. \end{cases}$$

Given p-efficient points $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k$,

$$\operatorname{conv}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k) + \mathbb{R}^n_+ \qquad \text{is an inner approximation of } \mathcal{L}.$$

Solution metods

Prékopa, Vizvári, Badics (1998):
set of p-efficient points generated before optimization,

Dentcheva, Prékopa, Ruszczyński (2000):
primal-dual method (cone-generation),

Dentcheva, Lai, Ruszczyński (2004), Dentcheva and Martinez (2013).

Cone generation Dentcheva, Prékopa, Ruszczyński (2000)

Primal problem formulated by splitting variables:

$$(\mathcal{P}) \qquad \min\ \boldsymbol{c}^T\boldsymbol{x} \quad \text{subject to} \quad T\boldsymbol{x} = \boldsymbol{z}, \quad \boldsymbol{x} \in X, \quad \boldsymbol{z} \in \mathcal{L}.$$

Lagrangian dual problem obtained by relaxing the constraint $T\boldsymbol{x} = \boldsymbol{z}$:

$$(\mathcal{D}) \qquad \max_{\boldsymbol{u}}\ D(\boldsymbol{u}) \quad \text{where} \quad D(\boldsymbol{u}) = \min_{\boldsymbol{x} \in X} (\boldsymbol{c}^T - \boldsymbol{u}^T T)\boldsymbol{x} \ + \ \min_{\boldsymbol{z} \in \mathcal{L}} \boldsymbol{u}^T \boldsymbol{z}.$$

Cone generation Dentcheva, Prékopa, Ruszczyński (2000)

Primal problem formulated by splitting variables:

$$(\mathcal{P}) \qquad \min \; \boldsymbol{c}^T \boldsymbol{x} \quad \text{subject to} \quad T\boldsymbol{x} = \boldsymbol{z}, \quad \boldsymbol{x} \in X, \quad \boldsymbol{z} \in \mathcal{L}.$$

Lagrangian dual problem obtained by relaxing the constraint $T\boldsymbol{x} = \boldsymbol{z}$:

$$(\mathcal{D}) \qquad \max_{\boldsymbol{u}} \; D(\boldsymbol{u}) \quad \text{where} \quad D(\boldsymbol{u}) = \min_{\boldsymbol{x} \in X} (\boldsymbol{c}^T - \boldsymbol{u}^T T)\boldsymbol{x} \; + \; \min_{\boldsymbol{z} \in \mathcal{L}} \boldsymbol{u}^T \boldsymbol{z}.$$

Solution method:

– from dual viewpoint: cutting-plane method for $D()$,

– from primal viewpoint: column generation method.

Cone generation Dentcheva, Prékopa, Ruszczyński (2000)

Primal problem formulated by splitting variables:

$$(\mathcal{P}) \qquad \min\ \boldsymbol{c}^T\boldsymbol{x} \quad \text{subject to} \quad T\boldsymbol{x} = \boldsymbol{z}, \quad \boldsymbol{x} \in X, \quad \boldsymbol{z} \in \mathcal{L}.$$

Lagrangian dual problem obtained by relaxing the constraint $T\boldsymbol{x} = \boldsymbol{z}$:

$$(\mathcal{D}) \qquad \max_{\boldsymbol{u}}\ D(\boldsymbol{u}) \quad \text{where} \quad D(\boldsymbol{u}) = \min_{\boldsymbol{x} \in X} (\boldsymbol{c}^T - \boldsymbol{u}^T T)\boldsymbol{x} \ + \ \min_{\boldsymbol{z} \in \mathcal{L}} \boldsymbol{u}^T \boldsymbol{z}.$$

Solution method:

– from dual viewpoint: cutting-plane method for $D()$,

– from primal viewpoint: column generation method.

New cuts/columns are improving p-efficient points.

These are found by solving subproblems $\min_{\boldsymbol{z} \in \mathcal{L}} \boldsymbol{u}^T \boldsymbol{z}$.

An alternative dual approach: epi-approximation

Idea: to build an inner approximation of the epigraph of the constraint function (instead of the level set $\mathcal{L}$).

An alternative dual approach: epi-approximation

Idea: to build an inner approximation of the epigraph of the constraint function (instead of the level set $\mathcal{L}$).

Constraint: $$F(z) \geq p$$

An alternative dual approach: epi-approximation

Idea: to build an inner approximation of the epigraph of the constraint function (instead of the level set $\mathcal{L}$).

Constraint: $F(z) \geq p$

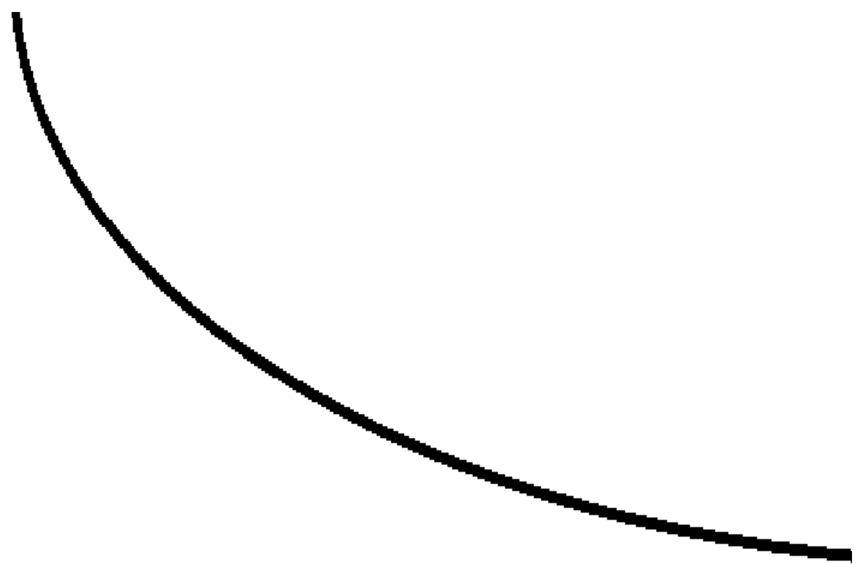
Convex formulation: $-\ln F(z) \leq -\ln p$

An alternative dual approach: epi-approximation

Idea: to build an inner approximation of the epigraph of the constraint function (instead of the level set $\mathcal{L}$).

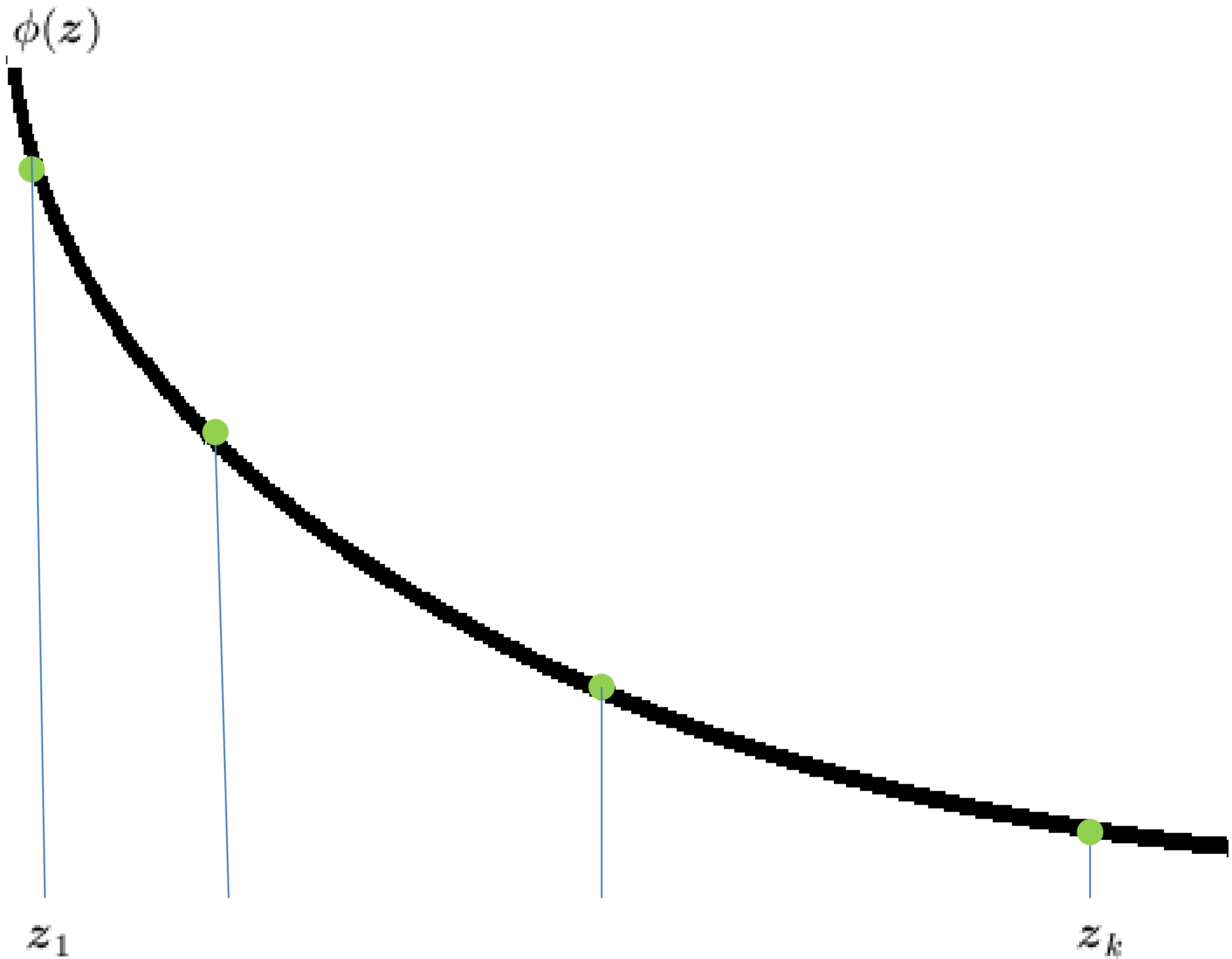
Constraint: $$F(z) \geq p$$

Convex formulation: $$\underbrace{-\ln F(z)}_{\phi(z)} \leq -\ln p$$



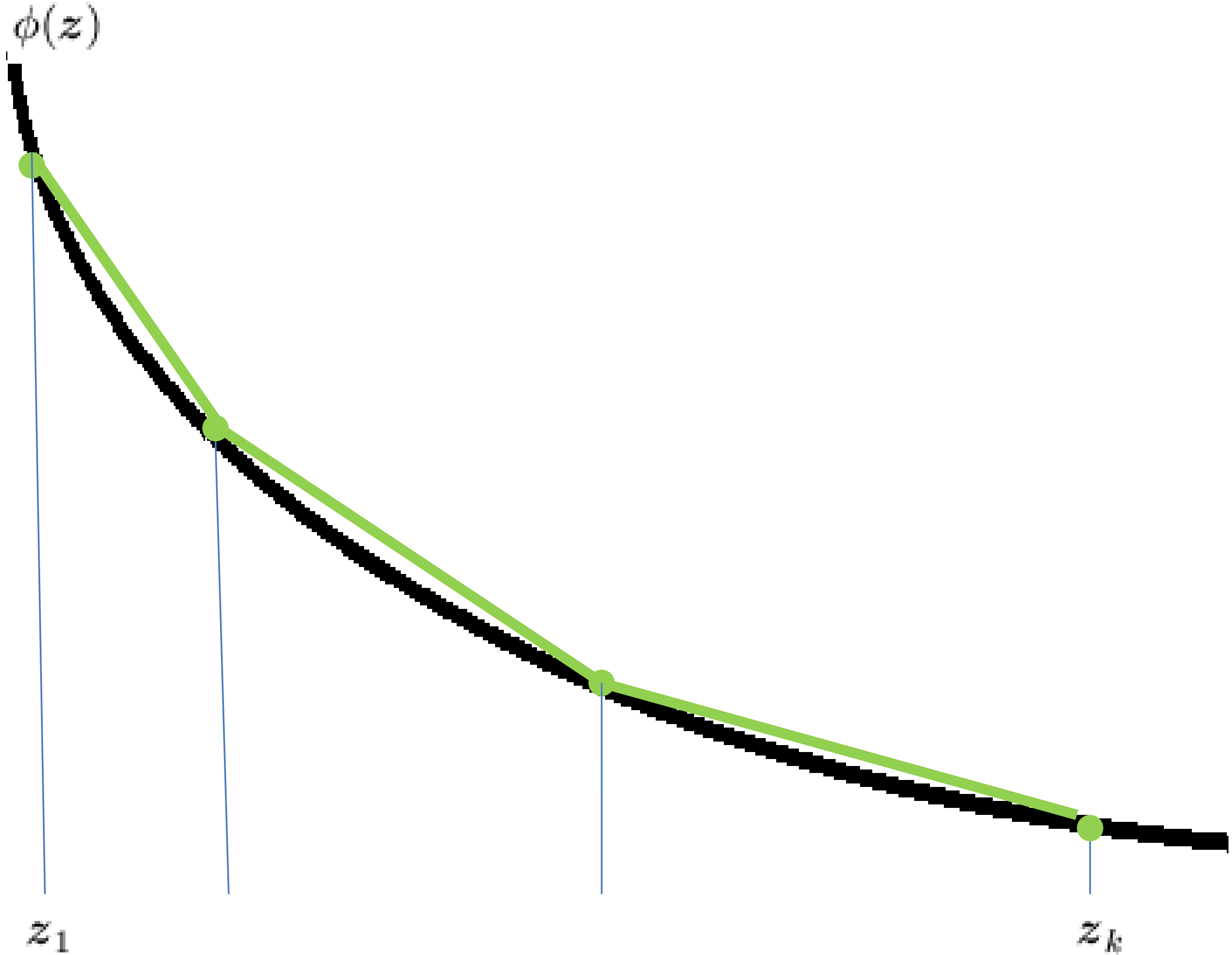
An alternative dual approach: epi-approximation

Given points $z_1, \ldots, z_k$, let $\phi_1 = \phi(z_1), \ldots, \phi_k = \phi(z_k)$.



An alternative dual approach: epi-approximation

Given points $z_1, \ldots, z_k$, let $\phi_1 = \phi(z_1), \ \ldots, \ \phi_k = \phi(z_k)$.



Epi-approximation to probability maximization

Primal problem formulated by splitting variables

$$(\mathcal{P}) \qquad \min\ \phi(\boldsymbol{z}) = -\ln F(\boldsymbol{z}) \quad \text{subject to} \quad T\boldsymbol{x} = \boldsymbol{z}, \quad \boldsymbol{x} \in X.$$

Lagrangian dual problem obtained by relaxing the constraint $T\boldsymbol{x} = \boldsymbol{z}$:

$$(\mathcal{D}) \qquad \max_{\boldsymbol{u}}\ D(\boldsymbol{u}) \quad \text{where} \quad D(\boldsymbol{u}) = \min_{\boldsymbol{x}\in X} \boldsymbol{u}^T T\boldsymbol{x} \ + \ \min_{\boldsymbol{z}} \left\{ \phi(\boldsymbol{z}) - \boldsymbol{u}^T \boldsymbol{z} \right\}.$$

Epi-approximation to probability maximization

Primal problem formulated by splitting variables

$$(\mathcal{P}) \qquad \min \; \phi(\boldsymbol{z}) = -\ln F(\boldsymbol{z}) \quad \text{subject to} \quad T\boldsymbol{x} = \boldsymbol{z}, \quad \boldsymbol{x} \in X.$$

Lagrangian dual problem obtained by relaxing the constraint $T\boldsymbol{x} = \boldsymbol{z}$:

$$(\mathcal{D}) \qquad \max_{\boldsymbol{u}} \; D(\boldsymbol{u}) \quad \text{where} \quad D(\boldsymbol{u}) = \min_{\boldsymbol{x}\in X} \boldsymbol{u}^T T\boldsymbol{x} \; + \; \min_{\boldsymbol{z}} \left\{ \phi(\boldsymbol{z}) - \boldsymbol{u}^T \boldsymbol{z} \right\}.$$

Solution method:

– from dual viewpoint: cutting-plane method for $D()$,

– from primal viewpoint: column generation method.

New cuts/columns are found by solving subproblems $\min_{\boldsymbol{z}} \left\{ \phi(\boldsymbol{z}) - \boldsymbol{u}^T \boldsymbol{z} \right\}$.

Epi-approximation computational study

Problems

Normal distribution, dimension up to 15.

Gradients of the distribution function can be computed componentwise:

$$\frac{\partial F(z_1,\ldots,z_n)}{\partial z_i} = F(z_1,\ldots,z_{i-1},z_{i+1},\ldots,z_n \mid z_i)\, f_i(z_i) \qquad (i = 1,\ldots,n).$$

Epi-approximation computational study

Problems

Normal distribution, dimension up to 15.

Gradients of the distribution function can be computed componentwise:

$$\frac{\partial F(z_1,\ldots,z_n)}{\partial z_i} = F(z_1,\ldots,z_{i-1},z_{i+1},\ldots,z_n \,|\, z_i)\; f_i(z_i) \qquad (i=1,\ldots,n).$$

Implementation: column generation scheme

Master problem solved by CPLEX simplex, version 12.6.3.

Subproblems solved by a steepest descent method.

$$\min_{\boldsymbol{z}} \left\{ \phi(\boldsymbol{z}) - \boldsymbol{u}^T \boldsymbol{z} \right\} \quad \text{where} \quad \phi(\boldsymbol{z}) = -\log F(\boldsymbol{z}).$$

$F(\boldsymbol{z})$ function values and gradients computed by Genz's code.
A gradient component can be computed from
an appropriate (n-1)-dimensional normal distribution function value.

Experience

Code proved reliable and robust.

Number of iterations depends on problem dimension, and on the optimal probability achieveable.

Almost all the computational effort was spent in the Genz subroutine !

Experience

Code proved reliable and robust.

Number of iterations depends on problem dimension,
and on the optimal probability achieveable.

Almost all the computational effort was spent in the Genz subroutine !

Heuristic adjustment

Aim: balancing different efforts:

– solving and resolving master problem (CPLEX),
– solving subproblems (Genz code).

Approximate solution of subproblems.

A single line search made in each steepest descent procedure,
and even this line search is approximate.

Experience

Code proved reliable and robust.
Number of iterations depends on problem dimension, and on the optimal probability achieveable.

Almost all the computational effort was spent in the Genz subroutine !

Heuristic adjustment

Aim: balancing different efforts:
– solving and resolving master problem (CPLEX),
– solving subproblems (Genz code).

Approximate solution of subproblems.
A single line search made in each steepest descent procedure, and even this line search is approximate.

Outcome

Effort spent in a subproblem substantially decreased

Number of master iterations did not increase significantly.

Theoretical justification of heuristic adjustment

An ideal convex programming problem

$\mathcal{F} \; = \; \min_{\boldsymbol{z}} f(\boldsymbol{z})$

Assume the function $f(\boldsymbol{z})$ is twice continuously differentiable,

and there are $\alpha, \omega \in \mathbb{R}$ $(0 < \alpha \leq \omega)$ such that

$$\alpha I \; \preceq \; \nabla^2 f(\boldsymbol{z}) \; \preceq \; \omega I \qquad (\boldsymbol{z} \in \mathbb{R}^n).$$

Here the relation $U \preceq V$ means that $V - U$ is positive semidefinite.

Theoretical justification of heuristic adjustment

An ideal convex programming problem

$\mathcal{F} = \min_{\boldsymbol{z}} f(\boldsymbol{z})$

Assume the function $f(\boldsymbol{z})$ is twice continuously differentiable,

and there are $\alpha, \omega \in \mathbb{R}$ $(0 < \alpha \leq \omega)$ such that

$$\alpha I \preceq \nabla^2 f(\boldsymbol{z}) \preceq \omega I \qquad (\boldsymbol{z} \in \mathbb{R}^n).$$

Here the relation $U \preceq V$ means that $V - U$ is positive semidefinite.

A well-known convergence theorem

We minimize $f(\boldsymbol{z})$ using a steepest descent method.

Starting from $\boldsymbol{z}^0$, let $\boldsymbol{z}^1, \ldots, \boldsymbol{z}^j, \ldots$ denote the iterates.

Then we have

$$f\left(\boldsymbol{z}^j\right) - \mathcal{F} \leq \left(1 - \tfrac{\alpha}{\omega}\right)^j \left[\, f\left(\boldsymbol{z}^0\right) - \mathcal{F}\,\right].$$

convergence theorem

$$f\left(\boldsymbol{z}^j\right) - \mathcal{F} \;\leq\; \left(1 - \tfrac{\alpha}{\omega}\right)^j \left[\, f\left(\boldsymbol{z}^0\right) - \mathcal{F}\,\right]$$

Application to our column generation scheme

New columns are found by solving subproblems

$$\min_{\boldsymbol{z}} \;\left\{\, \phi(\boldsymbol{z}) - \boldsymbol{u}^T \boldsymbol{z} \;\; + \text{constant} \,\right\}.$$

convergence theorem

$$f\left(\boldsymbol{z}^j\right) - \mathcal{F} \;\le\; \left(1 - \tfrac{\alpha}{\omega}\right)^j \left[\, f\left(\boldsymbol{z}^0\right) - \mathcal{F} \,\right]$$

Application to our column generation scheme

New columns are found by solving subproblems

$$\mathcal{F} = \min_{\boldsymbol{z}} \left\{ \underbrace{\phi(\boldsymbol{z}) - \boldsymbol{u}^T \boldsymbol{z} \;\; + \text{constant}}_{f(\boldsymbol{z})} \right\}.$$

Corollary of the theorem

We can find a starting point $\boldsymbol{z}^0$ such that

$$\boldsymbol{u}^T \boldsymbol{z}^j - \phi(\boldsymbol{z}^j) \;\; - \text{constant} \;\ge\; \left[1 - \left(1 - \tfrac{\alpha}{\omega}\right)^j\right] (-\mathcal{F}).$$

holds with iterate $\boldsymbol{z}^j$ obtained by j line search iterations.

Corollary :

We can find a starting point $\boldsymbol{z}^0$ such that

$$\boldsymbol{u}^T \boldsymbol{z}^j - \phi(\boldsymbol{z}^j) \ \ - \text{constant} \ \geq \ \left[1 - \left(1 - \tfrac{\alpha}{\omega}\right)^j\right] (-\mathcal{F}).$$

Corollary :

We can find a starting point $\boldsymbol{z}^0$ such that

$$\underbrace{\boldsymbol{u}^T \boldsymbol{z}^j - \phi(\boldsymbol{z}^j) \;\; - \text{constant}}_{\text{reduced cost of column } \boldsymbol{z}^j} \geq \left[1 - \left(1 - \tfrac{\alpha}{\omega}\right)^j\right] \underbrace{(-\mathcal{F})}_{\text{largest possible reduced cost value}}.$$

Consider the master problem as a LP problem.

Corollary :

We can find a starting point $\boldsymbol{z}^0$ such that

$$\underbrace{\boldsymbol{u}^T\boldsymbol{z}^j - \phi(\boldsymbol{z}^j) \;\; - \text{constant}}_{\text{reduced cost of column } \boldsymbol{z}^j} \;\geq\; \left[1 - \left(1 - \tfrac{\alpha}{\omega}\right)^j\right] \underbrace{(-\mathcal{F})}_{\text{largest possible reduced cost value}}.$$

Consider the master problem as a LP problem.

Even for a moderately large j, the iterate $\boldsymbol{z}^j$ is
a fairly good improving column in the column generation scheme.

On α, ω

Requirement: $\alpha I \preceq \nabla^2 \phi(\boldsymbol{z}) \preceq \omega I$

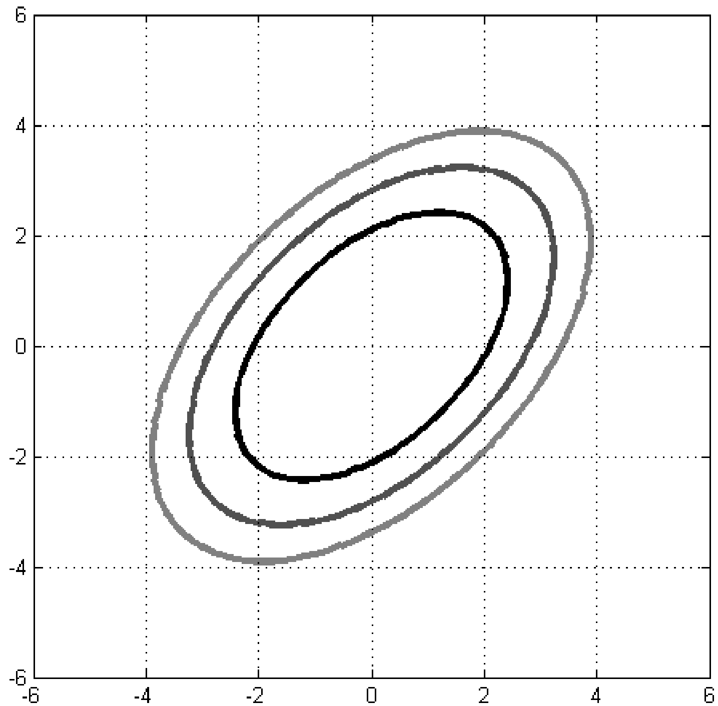
In the present model, we have $\phi(\boldsymbol{z}) = -\ln F(\boldsymbol{z})$.

Requirement does not hold for every $\boldsymbol{z}$.

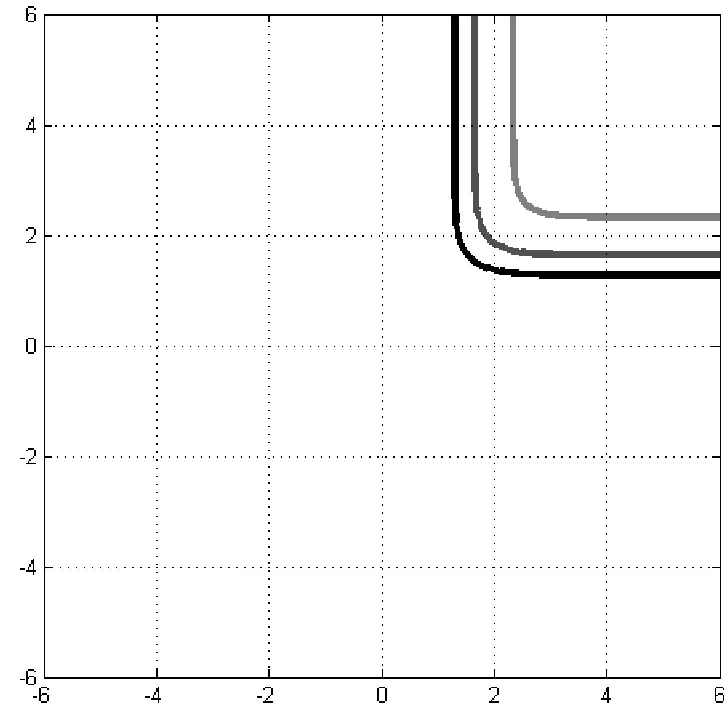
But it holds over a bounded box.

Illustration

Two-dimenisonal standard normal distribution, covariance = 0.5



Contours of density function



Contours of distribution function

Consider eigenvalues of $-\nabla^2 \log F(\boldsymbol{z})$.

Larger eigenvalue: $\omega(\boldsymbol{z})$.

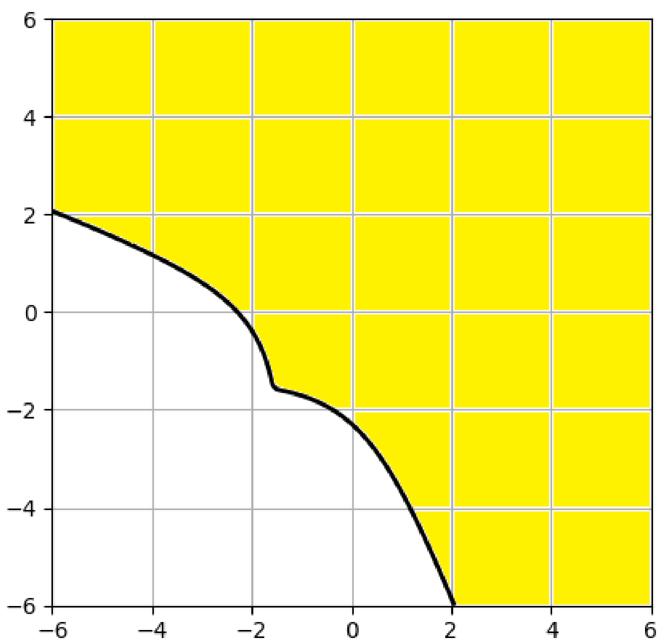
Smaller eigenvalue: $\alpha(\boldsymbol{z})$.

Consider eigenvalues of $-\nabla^2 \log F(\boldsymbol{z})$.

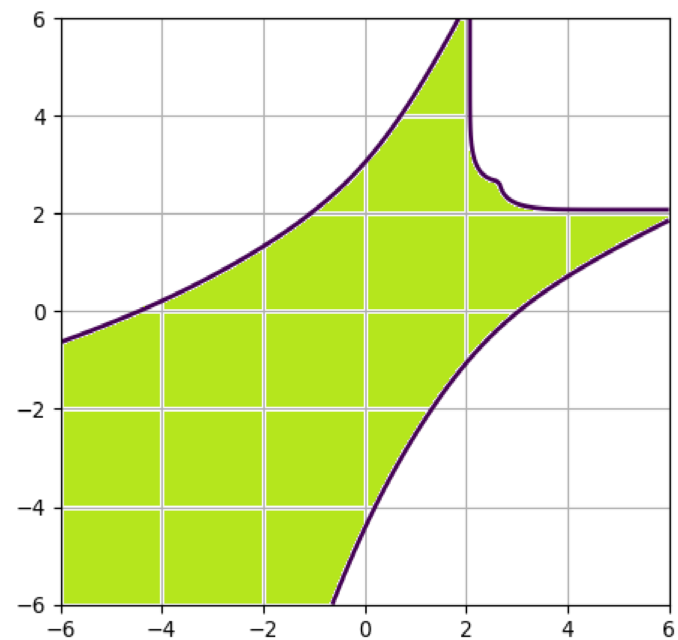
Larger eigenvalue: $\omega(\boldsymbol{z})$.

Smaller eigenvalue: $\alpha(\boldsymbol{z})$.

Level sets of eigenvalue functions



$\omega(\boldsymbol{z}) \le 10$



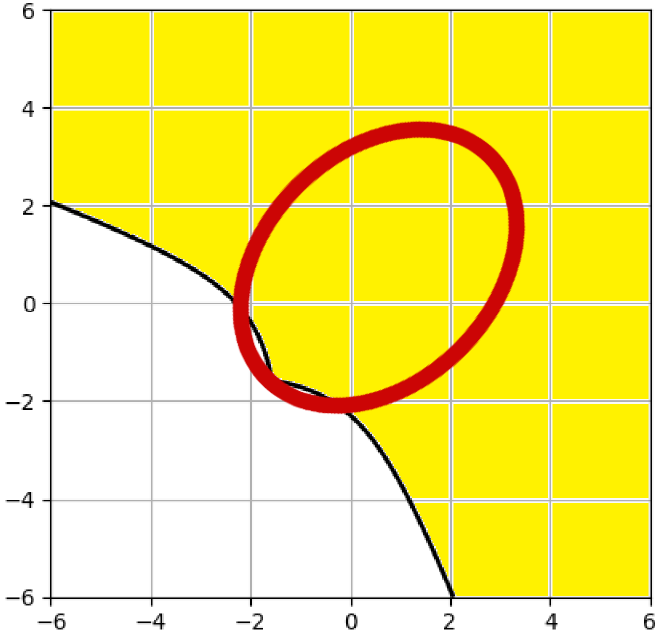
$\alpha(\boldsymbol{z}) \ge 0.1$

Consider eigenvalues of $-\nabla^2 \log F(\boldsymbol{z})$.

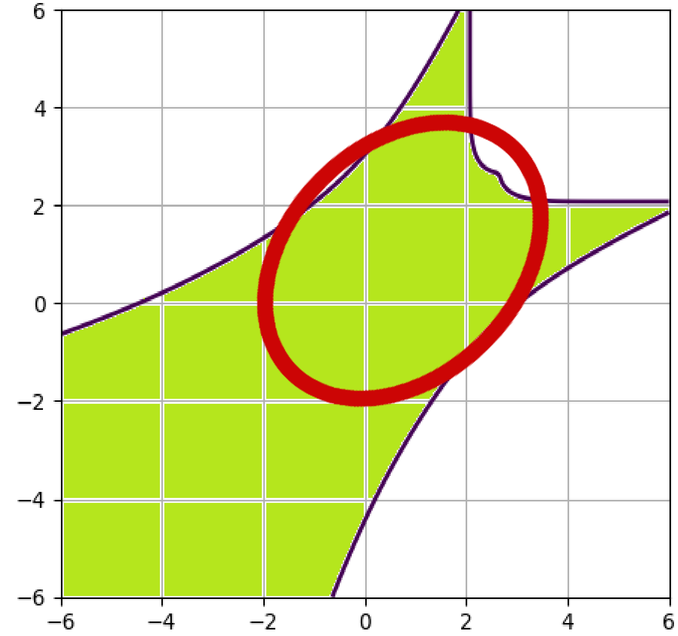
Larger eigenvalue: $\omega(\boldsymbol{z})$.

Smaller eigenvalue: $\alpha(\boldsymbol{z})$.

Level sets of eigenvalue functions



$\omega(\boldsymbol{z}) \leq 10$



$\alpha(\boldsymbol{z}) \geq 0.1$

Working with gradient estimates

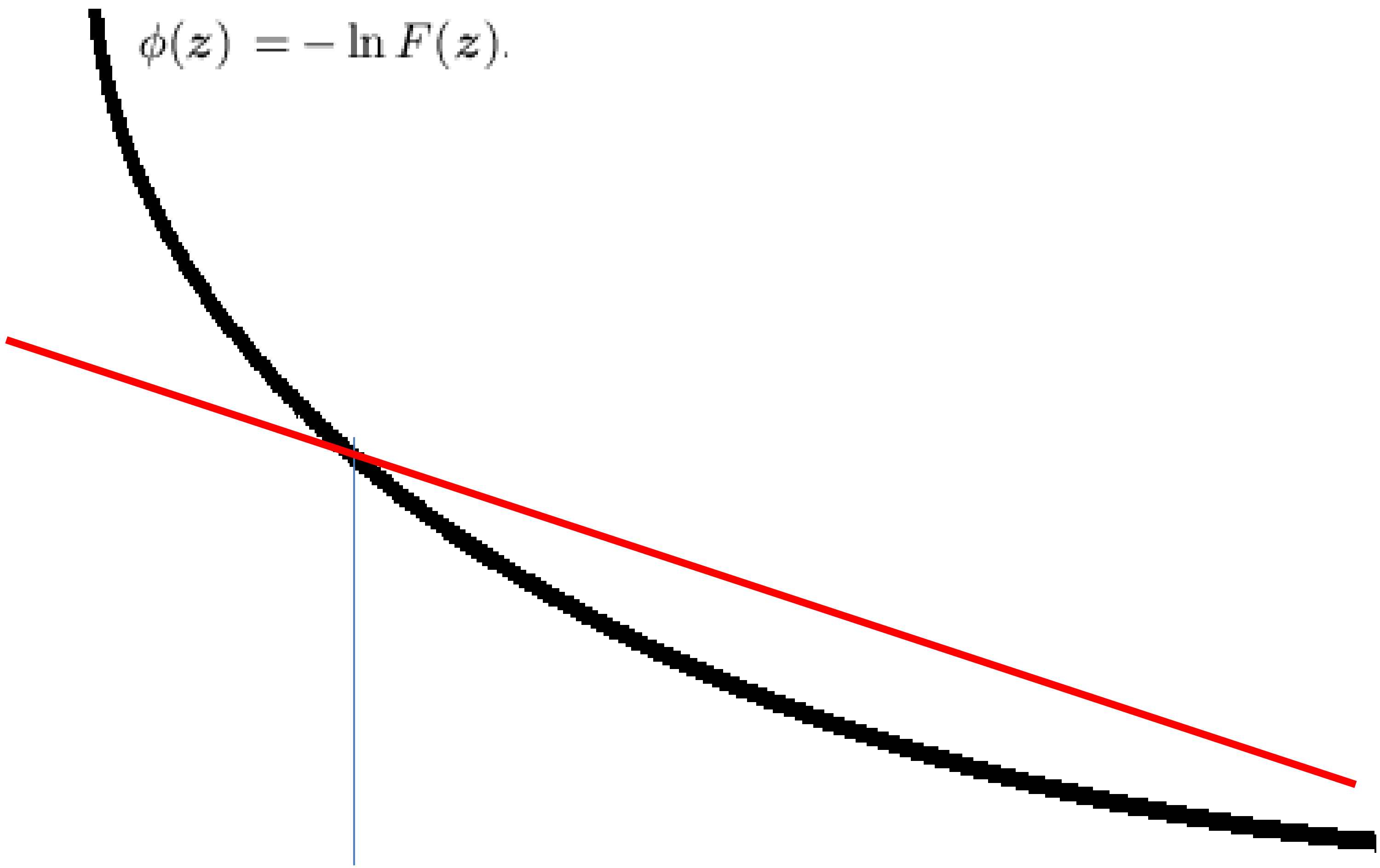
Feasibility study: comparison of direct cutting-plane method and dual approach

Working with gradient estimates

Feasibility study: comparison of direct cutting-plane method and dual approach

Direct cutting-plane method:

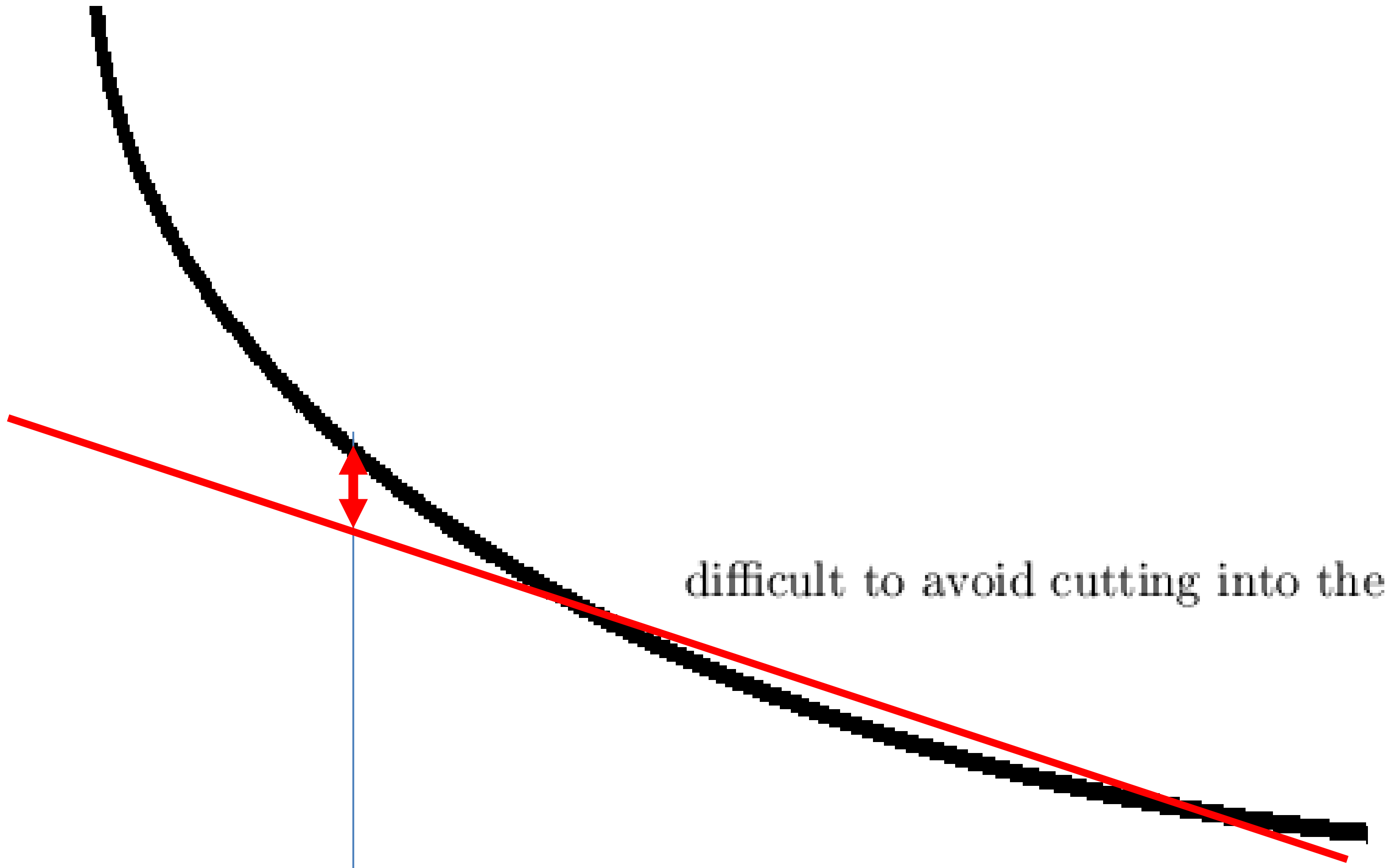
$$\phi(z) = -\ln F(z).$$



Working with gradient estimates

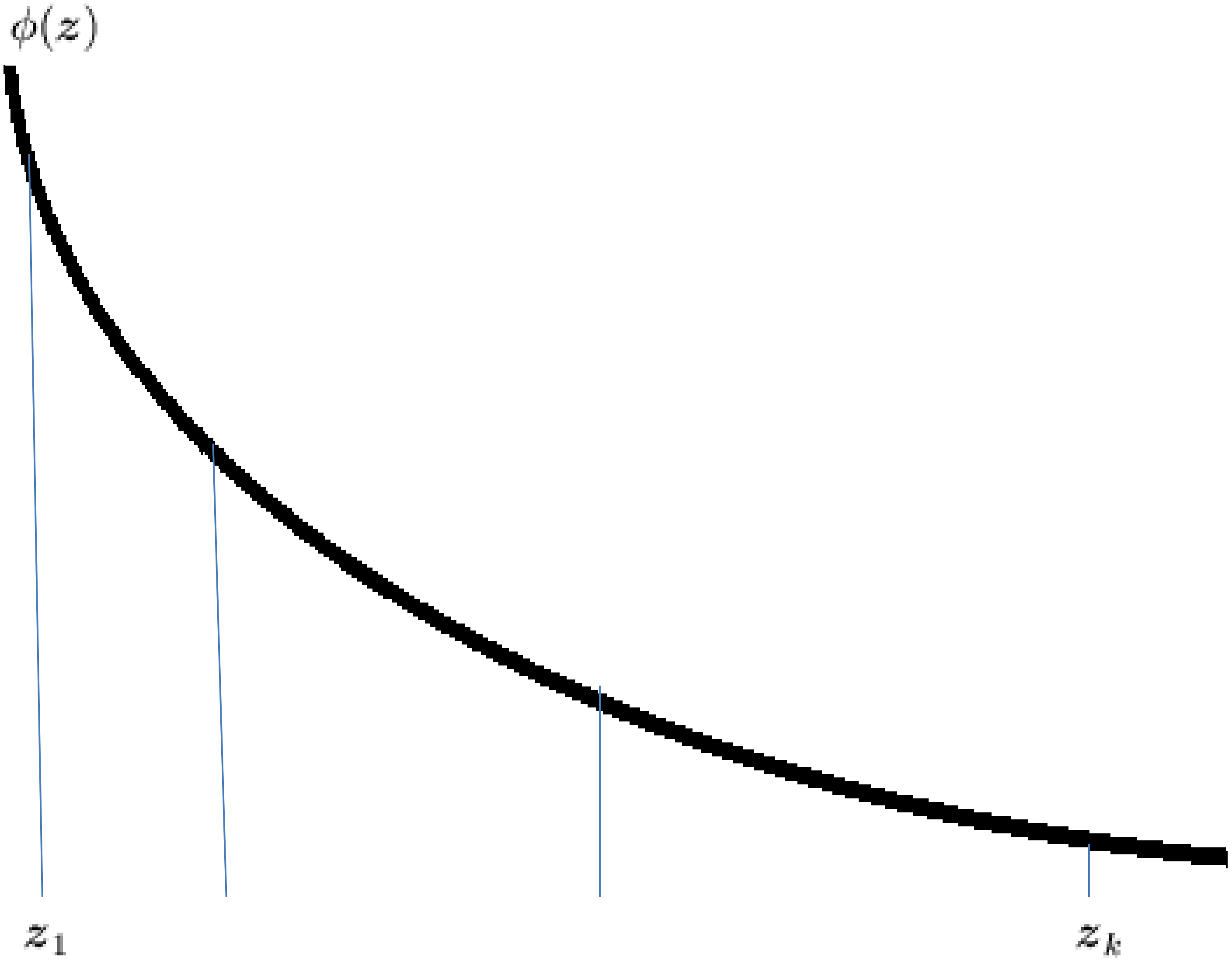
Feasibility study: comparison of direct cutting-plane method and dual approach

Direct cutting-plane method:

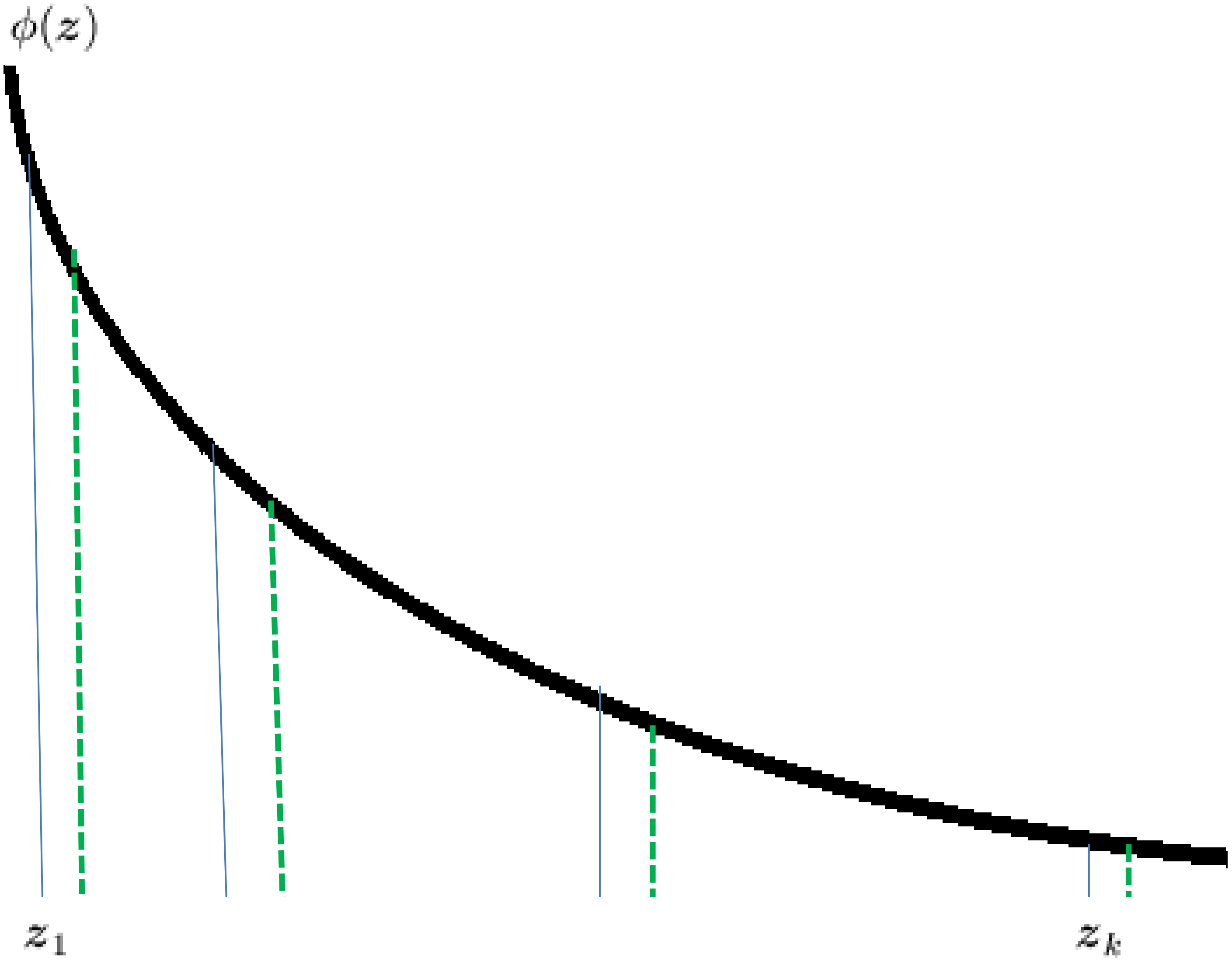


difficult to avoid cutting into the epigraph

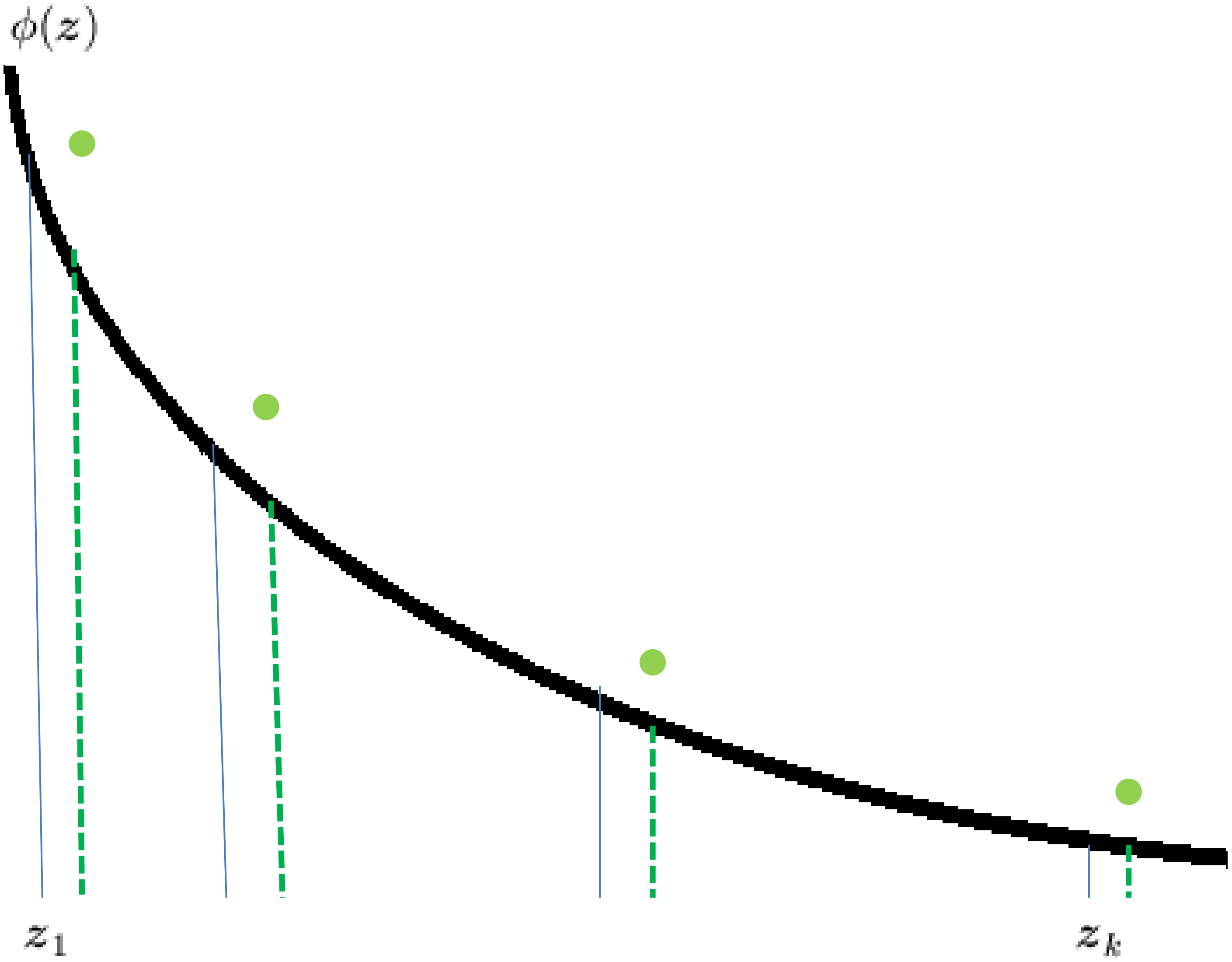
Epi-approximation



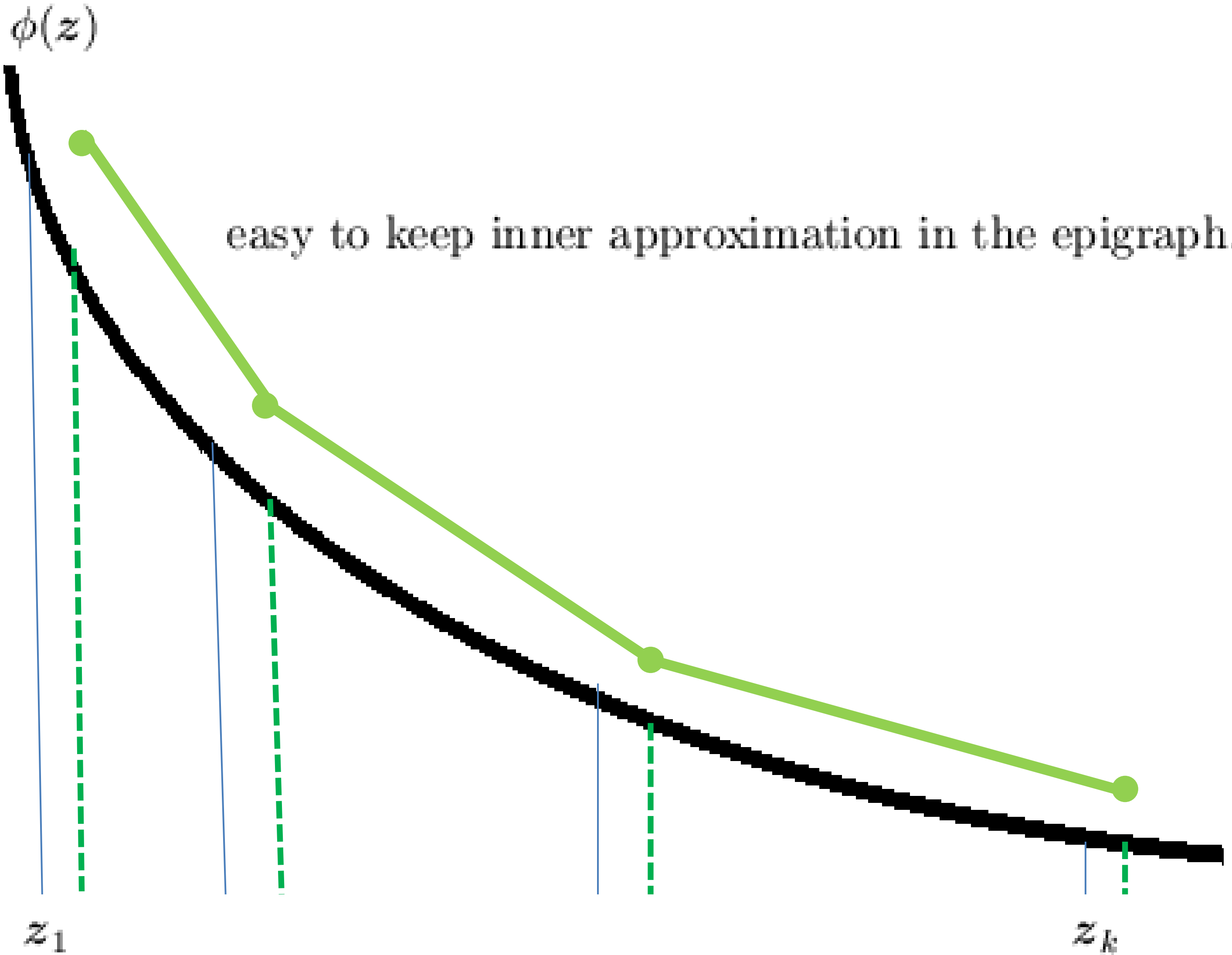
Epi-approximation



Epi-approximation



Epi-approximation



Working with gradient estimates

Theoretical development for an ideal convex programming problem

We wish to minimize $f(\boldsymbol{z})$ with a steepest descent method.

$$\alpha I \preceq \nabla^2 f(\boldsymbol{z}) \preceq \omega I \qquad (\boldsymbol{z} \in \mathbb{R}^n).$$

Working with gradient estimates

Theoretical development for an ideal convex programming problem

We wish to minimize $f(\boldsymbol{z})$ with a steepest descent method.

$$\alpha I \preceq \nabla^2 f(\boldsymbol{z}) \preceq \omega I \qquad (\boldsymbol{z} \in \mathbb{R}^n).$$

Given iterate $\boldsymbol{z}^\circ$, let $\boldsymbol{g}^\circ = -\nabla f(\boldsymbol{z}^\circ)$.

Computing $\boldsymbol{g}^\circ$ requires excessive effort.

Working with gradient estimates

Theoretical development for an ideal convex programming problem

We wish to minimize $f(\boldsymbol{z})$ with a steepest descent method.

$$\alpha I \ \preceq\ \nabla^2 f(\boldsymbol{z})\ \preceq\ \omega I \qquad (\boldsymbol{z} \in \mathbb{R}^n).$$

Given iterate $\boldsymbol{z}^\circ$, let $\boldsymbol{g}^\circ = -\nabla f(\boldsymbol{z}^\circ)$.

Computing $\boldsymbol{g}^\circ$ requires excessive effort.

Given $\sigma > 0$, we can construct realizations of a random vector $\boldsymbol{G}^\circ$ satisfying

$$\mathrm{E}\left(\boldsymbol{G}^\circ\right) = \boldsymbol{g}^\circ \qquad \text{and} \qquad \mathrm{E}\left(\left\|\boldsymbol{G}^\circ - \boldsymbol{g}^\circ\right\|^2\right)\ \leq\ \sigma \left\|\boldsymbol{g}^\circ\right\|^2.$$

Working with gradient estimates

Theoretical development for an ideal convex programming problem

Given $\sigma > 0$, we can construct realizations of a random vector $\boldsymbol{G}^{\circ}$ satisfying

$$\mathrm{E}\left(\boldsymbol{G}^{\circ}\right) = \boldsymbol{g}^{\circ} \qquad \text{and} \qquad \mathrm{E}\left(\left\|\boldsymbol{G}^{\circ} - \boldsymbol{g}^{\circ}\right\|^2\right) \le \sigma \left\|\boldsymbol{g}^{\circ}\right\|^2 .$$

A generalization of the convergence theorem

Starting from $\boldsymbol{z}^0$, let $\boldsymbol{z}^1, \ldots, \boldsymbol{z}^j, \ldots$ denote the iterates.

Then we have

$$\mathrm{E}\left[f\left(\boldsymbol{z}^j\right)\right] - \mathcal{F} \le \left(1 - \frac{\alpha}{\omega(\sigma+1)}\right)^j \left(f\left(\boldsymbol{z}^0\right) - \mathcal{F}\right).$$

Working with gradient estimates

Theoretical development for an ideal convex programming problem

Given $\sigma > 0$, we can construct realizations of a random vector $\boldsymbol{G}^{\circ}$ satisfying

$$\mathrm{E}\left(\boldsymbol{G}^{\circ}\right) = \boldsymbol{g}^{\circ} \qquad \text{and} \qquad \mathrm{E}\left(\left\|\boldsymbol{G}^{\circ} - \boldsymbol{g}^{\circ}\right\|^{2}\right) \leq \sigma \left\|\boldsymbol{g}^{\circ}\right\|^{2}.$$

A generalization of the convergence theorem

Starting from $\boldsymbol{z}^{0}$, let $\boldsymbol{z}^{1}, \ldots, \boldsymbol{z}^{j}, \ldots$ denote the iterates.

Then we have

$$\mathrm{E}\left[f\left(\boldsymbol{z}^{j}\right)\right] - \mathcal{F} \leq \left(1 - \tfrac{\alpha}{\omega(\sigma+1)}\right)^{j} \left(f\left(\boldsymbol{z}^{0}\right) - \mathcal{F}\right).$$

Resemblance to the stochastic approximation family

But present approach builds a model problem.

Working with gradient estimates

Application to probability maximization

Given $\sigma > 0$, we can construct realizations of a random vector $\boldsymbol{G}^\circ$ satisfying

$$\mathrm{E}\left(\boldsymbol{G}^\circ\right) = \boldsymbol{g}^\circ \qquad \text{and} \qquad \mathrm{E}\left(\left\|\boldsymbol{G}^\circ - \boldsymbol{g}^\circ\right\|^2\right) \le \sigma \left\|\boldsymbol{g}^\circ\right\|^2 .$$

Reliable gradient estimates can be constructed using ideas of

Szántai (1976, 1985);

Deák (1980, 1986);

Ambartzumian et al. (1998);

Gassmann (1988); Deák, Gassmann, Szántai (2002);

Mádi-Nagy, Prékopa (2004).

Köszönöm a figyelmet!