Nonlinear Derivative-free Constrained Optimization with a Mixed Penalty-Logarithmic Barrier Approach and Direct Search

Andrea Brilli[2], Ana Luísa Custódio[1], Giampaolo Liuzzi[2], and **Everton J Silva**[1]

[1]NOVA School of Science and Technology, NOVA MATH, Portugal
[2]"Sapienza" University of Rome, DIAG, Italy

UI/BD/151246/2021 UIDB/00297/2020 UIDP/00297/2020

Outline

Outline

Problem Features

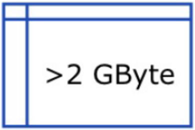
$$
\begin{aligned}
\min\ & f(\mathbf{x}) \\
\text{s.t.}\ & g(\mathbf{x}) \leq 0 \\
& h(\mathbf{x}) = 0 \\
& \mathbf{x} \in X
\end{aligned}
$$

$$
\begin{aligned}
f &: X \subseteq \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \\
g &: X \subseteq \mathbb{R}^n \to \{\mathbb{R} \cup \{+\infty\}\}^m \\
h &: X \subseteq \mathbb{R}^n \to \{\mathbb{R} \cup \{+\infty\}\}^p \\
X &= \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \leq \mathbf{b}\}
\end{aligned}
$$

- f, g, and h are black-box type and continuously differentiable
- Computing f, g, and h is expensive

No derivatives available for use

Long runtime



Large memory requirement

** Image credits to Joseph Simonis, ISMP 2009, Chicago, US*

SID-PSM: Direct Search using Simplex Derivatives

- Search step based on the minimization of some quadratic polynomial model (interpolation, minimum $\|.\|_F$ or regression)

  A. L. Custódio, H. Rocha, and L. N. Vicente
  Incorporating minimum Frobenius norm models in direct search
  Comput. Optim. Appl., 46: 265-278, 2010.

- Order of the poll vectors according to a negative simplex gradient

  A. L. Custódio and L. N. Vicente
  Using sampling and simplex derivatives in pattern search methods
  SIAM J. Optim., 18: 537-555, 2007.

- Handles constraints using an extreme barrier approach
- Efficient approaches to address general constraints (not yet!...)

SID-PSM: Direct Search using Simplex Derivatives

- Search step based on the minimization of some quadratic polynomial model (interpolation, minimum $\|.\|_F$ or regression)

  A. L. Custódio, H. Rocha, and L. N. Vicente
  Incorporating minimum Frobenius norm models in direct search
  Comput. Optim. Appl., 46: 265-278, 2010.

- Order of the poll vectors according to a negative simplex gradient

  A. L. Custódio and L. N. Vicente
  Using sampling and simplex derivatives in pattern search methods
  SIAM J. Optim., 18: 537-555, 2007.

- Handles constraints using an extreme barrier approach
- Efficient approaches to address general constraints (not yet!...)

SID-PSM: Direct Search using Simplex Derivatives

- Search step based on the minimization of some quadratic polynomial model (interpolation, minimum $\|.\|_F$ or regression)

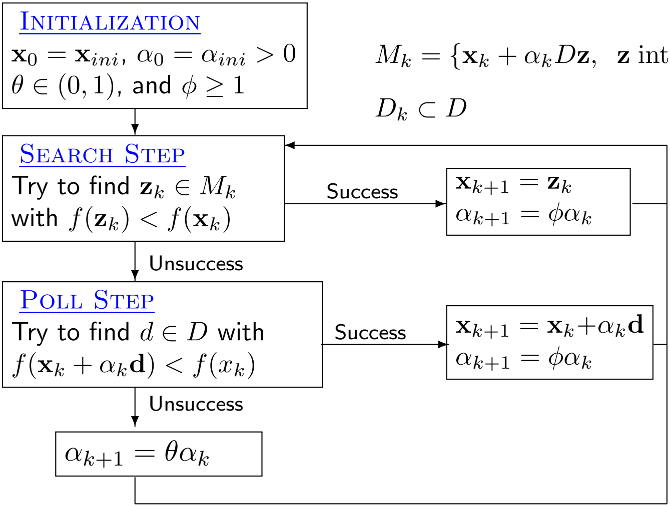
A. L. Custódio, H. Rocha, and L. N. Vicente
Incorporating minimum Frobenius norm models in direct search
Comput. Optim. Appl., 46: 265-278, 2010.

- Order of the poll vectors according to a negative simplex gradient

A. L. Custódio and L. N. Vicente
Using sampling and simplex derivatives in pattern search methods
SIAM J. Optim., 18: 537-555, 2007.

- Handles constraints using an extreme barrier approach
- Efficient approaches to address general constraints (not yet!...)

SID-PSM – Algorithmic Structure



$$M_k = \{\mathbf{x}_k + \alpha_k D\mathbf{z}, \ \ \mathbf{z} \text{ integer vector}\}$$

$$D_k \subset D$$

LOG-DFL: Logarithmic barrier penalty DFO algorithm

- New DFO method that uses a merit function with a log-barrier for inequality constraints and a penalty approach for equality constraints.

  A. Brilli, G. Liuzzi, and S. Lucidi
  An interior point method for nonlinear constrained derivative-free optimization
  arXiv, 2108.05157 [math.OC], 2021.

Mixed Penalty-Log Barrier Approach in Direct Search

Given an initial point $\mathbf{x}_0 \in X$, define:

$$\begin{aligned} \mathcal{G}^{\mathtt{log}} &= \{\ell \mid g_\ell(\mathbf{x}_0) < 0\} \\ \mathcal{G}^{\mathtt{ext}} &= \{\ell \mid g_\ell(\mathbf{x}_0) \geq 0\} \end{aligned}$$

Feasible region

$\mathcal{F} = X \cap \Omega_{\mathcal{G}^{\mathtt{log}}} \cap \Omega_{\mathcal{G}^{\mathtt{ext}}} \cap \Omega_h \neq \emptyset$, a compact set.

Mixed Penalty-Logarithmic Barrier function

$$Z(\mathbf{x};\rho) = f(\mathbf{x}) - \rho \sum_{\ell \in \mathcal{G}^{\mathtt{log}}} \log(-g_\ell(\mathbf{x})) + \frac{1}{\rho^{\nu-1}} \left(\sum_{\ell \in \mathcal{G}^{\mathtt{ext}}} (\max\{g_\ell(\mathbf{x}), 0\})^\nu + \sum_{j=1}^{p} |h_j(\mathbf{x})|^\nu \right),$$

where $\rho > 0$ and $\nu \in (1, 2]$.

Mixed Penalty-Log Barrier Approach in Direct Search

Given an initial point $\mathbf{x}_0 \in X$, define:

$$\begin{aligned} \mathcal{G}^{\texttt{log}} &= \{\ell \mid g_\ell(\mathbf{x}_0) < 0\} \\ \mathcal{G}^{\texttt{ext}} &= \{\ell \mid g_\ell(\mathbf{x}_0) \geq 0\} \end{aligned}$$

Feasible region

$\mathcal{F} = X \cap \Omega_{\mathcal{G}^{\texttt{log}}} \cap \Omega_{\mathcal{G}^{\texttt{ext}}} \cap \Omega_h \neq \emptyset$, a compact set.

Mixed Penalty-Logarithmic Barrier function

$$Z(\mathbf{x};\rho) = f(\mathbf{x}) - \rho \sum_{\ell \in \mathcal{G}^{\texttt{log}}} \log(-g_\ell(\mathbf{x})) + \frac{1}{\rho^{\nu-1}} \left(\sum_{\ell \in \mathcal{G}^{\texttt{ext}}} (\max\{g_\ell(\mathbf{x}), 0\})^\nu + \sum_{j=1}^{p} |h_j(\mathbf{x})|^\nu \right),$$

where $\rho > 0$ and $\nu \in (1, 2]$.

Mixed Penalty-Log Barrier Approach in Direct Search

Given an initial point $\mathbf{x}_0 \in X$, define:

$$\begin{aligned} \mathcal{G}^{\texttt{log}} &= \{\ell \mid g_\ell(\mathbf{x}_0) < 0\} \\ \mathcal{G}^{\texttt{ext}} &= \{\ell \mid g_\ell(\mathbf{x}_0) \geq 0\} \end{aligned}$$

Feasible region

$\mathcal{F} = X \cap \Omega_{\mathcal{G}^{\texttt{log}}} \cap \Omega_{\mathcal{G}^{\texttt{ext}}} \cap \Omega_h \neq \emptyset$, a compact set.

Mixed Penalty-Logarithmic Barrier function

$$Z(\mathbf{x};\rho) = f(\mathbf{x}) - \rho \sum_{\ell \in \mathcal{G}^{\texttt{log}}} \log(-g_\ell(\mathbf{x})) + \frac{1}{\rho^{\nu-1}} \left(\sum_{\ell \in \mathcal{G}^{\texttt{ext}}} (\max\{g_\ell(\mathbf{x}), 0\})^\nu + \sum_{j=1}^{p} |h_j(\mathbf{x})|^\nu \right),$$

where $\rho > 0$ and $\nu \in (1, 2]$.

Mixed Penalty-Logarithmic Barrier function

$$Z(\mathbf{x};\rho) = f(\mathbf{x}) - \rho \sum_{\ell \in \mathcal{G}^{\mathtt{log}}} \log(-g_\ell(\mathbf{x})) + \frac{1}{\rho^{\nu-1}} \left(\sum_{\ell \in \mathcal{G}^{\mathtt{ext}}} (\max\{g_\ell(\mathbf{x}), 0\})^\nu + \sum_{j=1}^{p} |h_j(\mathbf{x})|^\nu \right),$$

where $\rho > 0$ and $\nu \in (1, 2]$.

Penalized problem

$$\min \; Z(\mathbf{x};\rho)$$
$$\text{s.t. } \mathbf{x} \in X \cap \mathring{\Omega}_{\mathcal{G}^{\mathtt{log}}}$$

Equivalent Reformulation

$$\min \; Z(\mathbf{x};\rho)$$
$$\text{s.t. } \mathbf{x} \in X$$

- ρ must be driven to zero to solve the original problem

Mixed Penalty-Logarithmic Barrier function

$$Z(\mathbf{x};\rho) = f(\mathbf{x}) - \rho \sum_{\ell \in \mathcal{G}^{\mathtt{log}}} \log(-g_\ell(\mathbf{x})) + \frac{1}{\rho^{\nu-1}} \left(\sum_{\ell \in \mathcal{G}^{\mathtt{ext}}} (\max\{g_\ell(\mathbf{x}), 0\})^\nu + \sum_{j=1}^{p} |h_j(\mathbf{x})|^\nu \right),$$

where $\rho > 0$ and $\nu \in (1, 2]$.

Penalized problem

$$\begin{aligned} &\min \; Z(\mathbf{x};\rho) \\ &\text{s.t. } \mathbf{x} \in X \cap \mathring{\Omega}_{\mathcal{G}^{\mathtt{log}}} \end{aligned}$$

Equivalent Reformulation

$$\begin{aligned} &\min \; Z(\mathbf{x};\rho) \\ &\text{s.t. } \mathbf{x} \in X \end{aligned}$$

- ρ must be driven to zero to solve the original problem

Mixed Penalty-Logarithmic Barrier function

$$Z(\mathbf{x};\rho) = f(\mathbf{x}) - \rho \sum_{\ell \in \mathcal{G}^{\mathtt{log}}} \log(-g_\ell(\mathbf{x})) + \frac{1}{\rho^{\nu-1}} \left(\sum_{\ell \in \mathcal{G}^{\mathtt{ext}}} (\max\{g_\ell(\mathbf{x}), 0\})^\nu + \sum_{j=1}^{p} |h_j(\mathbf{x})|^\nu \right),$$

where $\rho > 0$ and $\nu \in (1, 2]$.

Penalized problem

$$\begin{aligned} &\min \ Z(\mathbf{x};\rho) \\ &\text{s.t. } \mathbf{x} \in X \cap \mathring{\Omega}_{\mathcal{G}^{\mathtt{log}}} \end{aligned}$$

Equivalent Reformulation

$$\begin{aligned} &\min \ Z(\mathbf{x};\rho) \\ &\text{s.t. } \mathbf{x} \in X \end{aligned}$$

- ρ must be driven to zero to solve the original problem

Outline

INITIALIZATION

$\mathbf{x}_0 \in X$ such that $g_\ell(\mathbf{x}_0) < 0$ for all $\ell \in \mathcal{G}^{\log}$, $\alpha_0 > 0$, $\mathcal{D}$, $\rho_0 > 0$, $\nu \in (1, 2]$, $\beta > 1$, $\theta_\alpha, \theta_\rho \in (0, 1)$, $\phi \geq 1$

$$\mathcal{D} = \bigcup_{k=0}^{+\infty} \mathbb{D}_k$$

STEP 1: SEARCH STEP

Try to find $\mathbf{z}_k \in X$ with

$Z(\mathbf{z}_k; \rho_k) \leq Z(\mathbf{x}_k; \rho_k) - \xi(\alpha_k)$

Success

$\mathbf{x}_{k+1} = \mathbf{z}_k$

$\alpha_{k+1} = \phi\alpha_k$

Unsuccess

STEP 2: POLL STEP

Try to find $d \in \mathbb{D}_k$ with

$Z(\mathbf{x}_k + \alpha_k d; \rho_k) \leq Z(\mathbf{x}_k; \rho_k) - \xi(\alpha_k)$

Success

$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k d$

$\alpha_{k+1} = \phi\alpha_k$

Unsuccess

$\alpha_{k+1} = \theta_\alpha \alpha_k$

$\alpha_{k+1} > \min\{\rho_k^\beta, (g_{\min})_k^2\}$

$(g_{\min})_k = \min_{\ell \in \mathcal{G}^{\log}} \{|g_\ell(\mathbf{x}_k)|\}$

$\alpha_{k+1} \leq \min\{\rho_k^\beta, (g_{\min})_k^2\}$

STEP 3: PENALTY UPDATE

$\rho_{k+1} = \theta_\rho \rho_k$

Penalty Parameter Update

Penalty Parameter Update.

Step 3. Set $(g_{\min})_k = \min_{\ell \in \mathcal{G}^{\texttt{log}}} \{|g_\ell(\mathbf{x}_{k+1})|\}$

If $\alpha_{k+1} < \alpha_k$ and $\alpha_{k+1} \leq \min\{\rho_k^\beta, (g_{\min})_k^2\}$

Then set $\rho_{k+1} = \theta_\rho \rho_k$

Else set $\rho_{k+1} = \rho_k$

The measure of stationarity α_k should go to zero faster than:

- the measure of quality ρ_k
- the measure of proximity $(g_{\min})_k$

Theorem

Let $\{\rho_k\}_{k\in\mathbb{N}}$ and $\{\alpha_k\}_{k\in\mathbb{N}}$ be sequences generated by LOG-SID-PSM. Then,

$$\lim_{k\to+\infty} \rho_k = 0 \text{ and } \lim_{k\to+\infty} \alpha_k = 0.$$

Penalty Parameter Update

Penalty Parameter Update.

Step 3. Set $(g_{\min})_k = \min_{\ell \in \mathcal{G}^{\texttt{log}}} \{|g_\ell(\mathbf{x}_{k+1})|\}$

If $\alpha_{k+1} < \alpha_k$ and $\alpha_{k+1} \leq \min\{\rho_k^\beta, (g_{\min})_k^2\}$

Then set $\rho_{k+1} = \theta_\rho \rho_k$

Else set $\rho_{k+1} = \rho_k$

The measure of stationarity α_k should go to zero faster than:

- the measure of quality ρ_k
- the measure of proximity $(g_{\min})_k$

Theorem

Let $\{\rho_k\}_{k\in\mathbb{N}}$ and $\{\alpha_k\}_{k\in\mathbb{N}}$ be sequences generated by LOG-SID-PSM. Then,

$$\lim_{k\to+\infty} \rho_k = 0 \text{ and } \lim_{k\to+\infty} \alpha_k = 0.$$

Penalty Parameter Update

Penalty Parameter Update.

Step 3. Set $(g_{\min})_k = \min\limits_{\ell \in \mathcal{G}^{\texttt{log}}} \{|g_\ell(\mathbf{x}_{k+1})|\}$

If $\alpha_{k+1} < \alpha_k$ and $\alpha_{k+1} \leq \min\{\rho_k^\beta, (g_{\min})_k^2\}$

Then set $\rho_{k+1} = \theta_\rho \rho_k$

Else set $\rho_{k+1} = \rho_k$

The measure of stationarity α_k should go to zero faster than:

- the measure of quality ρ_k
- the measure of proximity $(g_{\min})_k$

Theorem

Let $\{\rho_k\}_{k\in\mathbb{N}}$ and $\{\alpha_k\}_{k\in\mathbb{N}}$ be sequences generated by LOG-SID-PSM. Then,

$$\lim_{k\to+\infty} \rho_k = 0 \text{ and } \lim_{k\to+\infty} \alpha_k = 0.$$

Penalty Parameter Update

Penalty Parameter Update.

Step 3. Set $(g_{\min})_k = \min_{\ell \in \mathcal{G}^{\texttt{log}}} \{|g_\ell(\mathbf{x}_{k+1})|\}$

If $\alpha_{k+1} < \alpha_k$ and $\alpha_{k+1} \leq \min\{\rho_k^\beta, (g_{\min})_k^2\}$

Then set $\rho_{k+1} = \theta_\rho \rho_k$

Else set $\rho_{k+1} = \rho_k$

The measure of stationarity α_k should go to zero faster than:

- the measure of quality ρ_k
- the measure of proximity $(g_{\min})_k$

Theorem

Let $\{\rho_k\}_{k\in\mathbb{N}}$ and $\{\alpha_k\}_{k\in\mathbb{N}}$ be sequences generated by LOG-SID-PSM. Then,

$$\lim_{k\to+\infty} \rho_k = 0 \text{ and } \lim_{k\to+\infty} \alpha_k = 0.$$

Penalty Parameter Update

Penalty Parameter Update.

Step 3. Set $(g_{\min})_k = \min_{\ell \in \mathcal{G}^{\texttt{log}}} \{|g_\ell(\mathbf{x}_{k+1})|\}$

If $\alpha_{k+1} < \alpha_k$ and $\alpha_{k+1} \leq \min\{\rho_k^\beta, (g_{\min})_k^2\}$

Then set $\rho_{k+1} = \theta_\rho \rho_k$

Else set $\rho_{k+1} = \rho_k$

The measure of stationarity α_k should go to zero faster than:

- the measure of quality ρ_k
- the measure of proximity $(g_{\min})_k$

Theorem

Let $\{\rho_k\}_{k\in\mathbb{N}}$ and $\{\alpha_k\}_{k\in\mathbb{N}}$ be sequences generated by LOG-SID-PSM. Then,

$$\lim_{k\to+\infty} \rho_k = 0 \text{ and } \lim_{k\to+\infty} \alpha_k = 0.$$

Penalty Parameter Update

Penalty Parameter Update.

Step 3. Set $(g_{\min})_k = \min_{\ell \in \mathcal{G}^{\texttt{log}}} \{|g_\ell(\mathbf{x}_{k+1})|\}$

If $\alpha_{k+1} < \alpha_k$ and $\alpha_{k+1} \leq \min\{\rho_k^\beta, (g_{\min})_k^2\}$

Then set $\rho_{k+1} = \theta_\rho \rho_k$

Else set $\rho_{k+1} = \rho_k$

The measure of stationarity α_k should go to zero faster than:

- the measure of quality ρ_k
- the measure of proximity $(g_{\min})_k$

Theorem

Let $\{\rho_k\}_{k\in\mathbb{N}}$ and $\{\alpha_k\}_{k\in\mathbb{N}}$ be sequences generated by LOG-SID-PSM. Then,

$$\lim_{k\to+\infty} \rho_k = 0 \text{ and } \lim_{k\to+\infty} \alpha_k = 0.$$

Globalization Strategy

Forcing Function $\xi : [0, +\infty) \to [0, +\infty)$

- a continuous and nondecreasing function
- $\xi(t)/t \to 0$ when $t \to 0$
- If $\xi(t) \to 0$ then $t \to 0$

Mangasarian-Fromovitz Type Constraint Qualification

MFCQ

Let $x \in X$ and $T_X(\mathbf{x})$ be the tangent cone at $\mathbf{x}$ with respect to the linear constraints. The point $\mathbf{x}$ is said to satisfy the MFCQ if the two following conditions are satisfied:

(a) There does not exist a nonzero vector $\alpha = (\alpha_1, \ldots, \alpha_q)$ such that:

$$\left(\sum_{i=1}^{q} \alpha_i \nabla h_i(\mathbf{x}) \right)^\top \mathbf{d} \geq 0, \qquad \forall \mathbf{d} \in T_X(\mathbf{x});$$

(b) There exists a feasible direction $\mathbf{d} \in T_X(\mathbf{x})$, such that:

$$\nabla g_\ell(\mathbf{x})^\top \mathbf{d} < 0, \quad \forall \ell \in I_+(\mathbf{x}), \quad \nabla h_j(\mathbf{x})^\top \mathbf{d} = 0, \quad \forall j = 1, \ldots, p$$

where $I_+(\mathbf{x}) = \{\ell \mid g_\ell(\mathbf{x}) \geq 0\}$.

Active Constraints and Tangent Cone

For every $\mathbf{x} \in X$, i.e., such that $\mathbf{A}\mathbf{x} \leq \mathbf{b}$:

$I_X(\mathbf{x}) = \{i \mid \mathbf{a}_i^\top \mathbf{x} = b_i\}$ (set of indices of active constraints)

$T_X(\mathbf{x}) = \{\mathbf{d} \in \mathbb{R}^n \mid \mathbf{a}_i^\top \mathbf{d} \leq 0,\ i \in I_X(\mathbf{x})\}$ (tangent cone at $\mathbf{x}$)

ε-Active Constraints and Tangent Cone

For every $\mathbf{x} \in X$, i.e., such that $\mathbf{Ax} \leq \mathbf{b}$:

$I_X(\mathbf{x}_k, \varepsilon) = \{i \mid \mathbf{a}_i^\top \mathbf{x}_k \geq b_i - \varepsilon\}$ (set of indices of ε-active constraints)

$T_X(\mathbf{x}_k, \varepsilon) = \{\mathbf{d} \in \mathbb{R}^n \mid \mathbf{a}_i^\top \mathbf{d} \leq 0,\ i \in I_X(\mathbf{x}_k, \varepsilon)\}$ (ε-tangent cone at $\mathbf{x}_k$)

Proposition

Let $\{\mathbf{x}_k\}_{k\in\mathbb{N}}$ be a sequence of points in X converging to $\mathbf{x}^* \in X$. Then, there exists an $\varepsilon^* > 0$ (depending only on $\mathbf{x}^*$) such that for any $\varepsilon \in (0, \varepsilon^*]$ there exists $k_\varepsilon \in \mathbb{N}$ such that

$$I_X(\mathbf{x}^*) = I_X(\mathbf{x}_k, \varepsilon)$$
$$T_X(\mathbf{x}^*) = T_X(\mathbf{x}_k, \varepsilon)$$

for all $k \geq k_\varepsilon$.

ε-Active Constraints and Tangent Cone

For every $\mathbf{x} \in X$, i.e., such that $\mathbf{A}\mathbf{x} \leq \mathbf{b}$:

$I_X(\mathbf{x}_k, \varepsilon) = \{i \mid \mathbf{a}_i^\top \mathbf{x}_k \geq b_i - \varepsilon\}$ (set of indices of ε-active constraints)

$T_X(\mathbf{x}_k, \varepsilon) = \{\mathbf{d} \in \mathbb{R}^n \mid \mathbf{a}_i^\top \mathbf{d} \leq 0,\ i \in I_X(\mathbf{x}_k, \varepsilon)\}$ (ε-tangent cone at $\mathbf{x}_k$)

Proposition

Let $\{\mathbf{x}_k\}_{k\in\mathbb{N}}$ be a sequence of points in X converging to $\mathbf{x}^* \in X$. Then, there exists an $\varepsilon^* > 0$ (depending only on $\mathbf{x}^*$) such that for any $\varepsilon \in (0, \varepsilon^*]$ there exists $k_\varepsilon \in \mathbb{N}$ such that

$$I_X(\mathbf{x}^*) = I_X(\mathbf{x}_k, \varepsilon)$$
$$T_X(\mathbf{x}^*) = T_X(\mathbf{x}_k, \varepsilon)$$

for all $k \geq k_\varepsilon$.

Geometry Assumption

Let $\{\mathbf{x}_k\}_{k\in\mathbb{N}}$ be a sequence of points such that $\mathbf{x}_k \in X$. The sequence $\mathbb{D}_k$ of poll directions satisfies:

$$\mathbb{D}_k = \{\mathbf{d}_k^i \mid \|\mathbf{d}_k^i\| = 1, i = 1, \ldots, |\mathbb{D}_k|\}$$

and for some $\bar{\varepsilon} > 0$,

$$\mathrm{cone}(\mathbb{D}_k \cap T_X(\mathbf{x}_k, \varepsilon)) = T_X(\mathbf{x}_k, \varepsilon), \quad \forall\, \varepsilon \in (0, \bar{\varepsilon}].$$

Furthermore, $\mathcal{D} = \bigcup_{k=0}^{+\infty} \mathbb{D}_k$ is a finite set, and $|\mathbb{D}_k|$ is bounded.

Convergence of LOG-SID-PSM

Lagrange Multipliers

$$\nabla Z(\mathbf{x};\rho_k) = \nabla f(\mathbf{x}) + \sum_{\ell \in \mathcal{G}^{\texttt{log}}} \frac{\rho_k}{-g_\ell(\mathbf{x})} \nabla g_\ell(\mathbf{x}) + \sum_{\ell \in \mathcal{G}^{\texttt{ext}}} \nu \left(\frac{\max\{g_\ell(\mathbf{x}), 0\}}{\rho_k} \right)^{\nu-1} \nabla g_\ell(\mathbf{x}) + \sum_{j=1}^{p} \nu \left(\frac{|h_j(\mathbf{x})|}{\rho_k} \right)^{\nu-1} \nabla h_j(\mathbf{x})$$

$$\lambda_\ell(\mathbf{x};\rho) = \begin{cases} \dfrac{\rho}{-g_\ell(\mathbf{x})}, & \text{if } \ell \in \mathcal{G}^{\texttt{log}} \\ \nu \left(\dfrac{\max\{g_\ell(\mathbf{x}), 0\}}{\rho} \right)^{\nu-1}, & \text{if } \ell \in \mathcal{G}^{\texttt{ext}} \end{cases}$$

$$\mu_j(\mathbf{x};\rho) = \nu \left(\frac{|h_j(\mathbf{x})|}{\rho} \right)^{\nu-1}, \quad j = 1, \ldots, p$$

Convergence of LOG-SID-PSM

Lagrange Multipliers

$$\nabla Z(\mathbf{x};\rho_k) = \nabla f(\mathbf{x}) + \sum_{\ell\in\mathcal{G}^{\texttt{log}}} \frac{\rho_k}{-g_\ell(\mathbf{x})}\nabla g_\ell(\mathbf{x}) + \sum_{\ell\in\mathcal{G}^{\texttt{ext}}} \nu\left(\frac{\max\{g_\ell(\mathbf{x}),0\}}{\rho_k}\right)^{\nu-1}\nabla g_\ell(\mathbf{x}) +$$

$$+\sum_{j=1}^{p}\nu\left(\frac{|h_j(\mathbf{x})|}{\rho_k}\right)^{\nu-1}\nabla h_j(\mathbf{x})$$

$$\lambda_\ell(\mathbf{x};\rho) = \begin{cases} \dfrac{\rho}{-g_\ell(\mathbf{x})}, & \text{if } \ell\in\mathcal{G}^{\texttt{log}} \\ \nu\left(\dfrac{\max\{g_\ell(\mathbf{x}),0\}}{\rho}\right)^{\nu-1}, & \text{if } \ell\in\mathcal{G}^{\texttt{ext}} \end{cases}$$

$$\mu_j(\mathbf{x};\rho) = \nu\left(\frac{|h_j(\mathbf{x})|}{\rho}\right)^{\nu-1}, \quad j=1,\ldots,p$$

Convergence of LOG-SID-PSM

Theorem

Let $\{\mathbf{x}_k\}_{k\in\mathbb{N}}$ be the sequence of iterates generated by LOG-SID-PSM. Consider the set $K = \{k \in \mathbb{N} : \rho_{k+1} < \rho_k\}$, assume that the sets of directions $\{\mathbb{D}_k\}_{k\in\mathbb{N}}$, used by the algorithm, satisfy the Geometry Assumption and let x^* be a limit point of $\{\mathbf{x}_k\}_{k\in\hat{K}}$, $\hat{K} \subseteq K$, that satisfies the MFCQ. Then

(i) The sequences of Lagrange multipliers $\{\lambda_\ell(\mathbf{x}_k;\rho_k)\}_{k\in\hat{K}}$, $\ell = 1,\ldots,m$ and $\{\mu_j(\mathbf{x}_k;\rho_k)\}_{k\in\hat{K}}$, $j = 1,\ldots,p$ are bounded.

(ii) x^* is a stationary point

Outline

LOG-SID-PSM

- Direct Search Matlab code
- Attempts a search step based on the minimization of quadratic models
- Orders the poll vectors according to a negative simplex gradient
- Handles constraints using a mixed Penalty-Logarithmic Barrier

Sufficient decrease condition

$$Z(\mathbf{x}_{k+1};\rho_k) \leq Z(\mathbf{x}_k;\rho_k) - \gamma\alpha_k^2,$$

where $\gamma = 10^{-9}$

LOG-SID-PSM

- Direct Search Matlab code
- Attempts a search step based on the minimization of quadratic models
- Orders the poll vectors according to a negative simplex gradient
- Handles constraints using a mixed Penalty-Logarithmic Barrier

Sufficient decrease condition

$$Z(\mathbf{x}_{k+1}; \rho_k) \leq Z(\mathbf{x}_k; \rho_k) - \gamma \alpha_k^2,$$

where $\gamma = 10^{-9}$

LOG-SID-PSM

- Direct Search Matlab code
- Attempts a search step based on the minimization of quadratic models
- Orders the poll vectors according to a negative simplex gradient
- Handles constraints using a mixed Penalty-Logarithmic Barrier

Sufficient decrease condition

$$Z(\mathbf{x}_{k+1}; \rho_k) \leq Z(\mathbf{x}_k; \rho_k) - \gamma \alpha_k^2,$$

where $\gamma = 10^{-9}$

LOG-SID-PSM

- Direct Search Matlab code
- Attempts a search step based on the minimization of quadratic models
- Orders the poll vectors according to a negative simplex gradient
- Handles constraints using a mixed Penalty-Logarithmic Barrier

Sufficient decrease condition

$$Z(\mathbf{x}_{k+1}; \rho_k) \leq Z(\mathbf{x}_k; \rho_k) - \gamma \alpha_k^2,$$

where $\gamma = 10^{-9}$

LOG-SID-PSM

- Direct Search Matlab code
- Attempts a search step based on the minimization of quadratic models
- Orders the poll vectors according to a negative simplex gradient
- Handles constraints using a mixed Penalty-Logarithmic Barrier

Sufficient decrease condition

$$Z(\mathbf{x}_{k+1};\rho_k) \leq Z(\mathbf{x}_k;\rho_k) - \gamma\alpha_k^2,$$

where $\gamma = 10^{-9}$

Search Step in LOG-SID-PSM – Model Building

- Reuses previous function evaluations and select points in $B(\mathbf{x}_k; \Delta_k)$, with

$$\Delta_k \;=\; \sigma \, \alpha_k \max_{\mathbf{d} \in D_{k-1}} \|\mathbf{d}\|$$

- Builds quadratic models for each function (f, g, and h) using the selected points:

  $\#$ points in $[n+2, (n+1)(n+2)/2[\Rightarrow$ MFN model

  $\#$ points $= (n+1)(n+2)/2 \Rightarrow$ Determined interpolation model

  $\#$ points in $](n+1)(n+2)/2, (n+1)(n+2)] \Rightarrow$ Regression model
  - 80% of the points selected nearest to the current iterate
  - 20% of the points selected farthest away from the current iterate

Search Step in LOG-SID-PSM – Model Building

- Reuses previous function evaluations and select points in $B(\mathbf{x}_k; \Delta_k)$, with

$$\Delta_k = \sigma \, \alpha_k \max_{\mathbf{d} \in D_{k-1}} \|\mathbf{d}\|$$

- Builds quadratic models for each function (f, g, and h) using the selected points:

points in $[n+2, (n+1)(n+2)/2[\Rightarrow$ MFN model

points $= (n+1)(n+2)/2 \Rightarrow$ Determined interpolation model

points in $](n+1)(n+2)/2, (n+1)(n+2)] \Rightarrow$ Regression model

  - 80% of the points selected nearest to the current iterate
  - 20% of the points selected farthest away from the current iterate

Search Step in LOG-SID-PSM – Model Building

- Reuses previous function evaluations and select points in $B(\mathbf{x}_k; \Delta_k)$, with

$$\Delta_k \;=\; \sigma \, \alpha_k \; \max_{\mathbf{d} \in D_{k-1}} \|\mathbf{d}\|$$

- Builds quadratic models for each function (f, g, and h) using the selected points:

  $\#$ points in $[n+2, (n+1)(n+2)/2[\Rightarrow$ MFN model

  $\#$ points $= (n+1)(n+2)/2 \;\Rightarrow$ Determined interpolation model

  $\#$ points in $](n+1)(n+2)/2, (n+1)(n+2)] \;\Rightarrow$ Regression model
  - 80% of the points selected nearest to the current iterate
  - 20% of the points selected farthest away from the current iterate

Search Step in LOG-SID-PSM

- Computes $\mathbf{z}_k$ as the solution to the following problem:

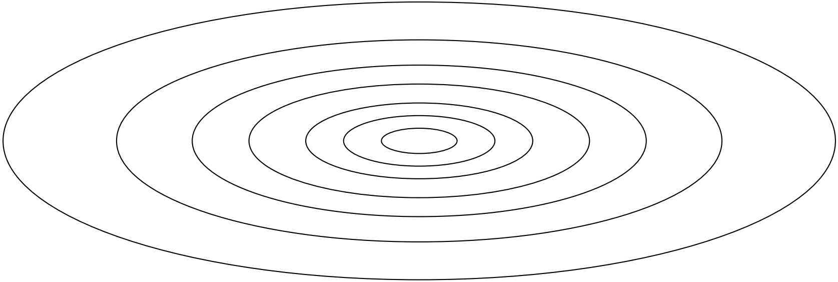
$$\begin{array}{l} \min \;\; Z_m(\mathbf{z};\rho_k) \\ \text{s.t. } \mathbf{z} \in X \cap B(x_k;\Delta_k) \end{array}$$

where

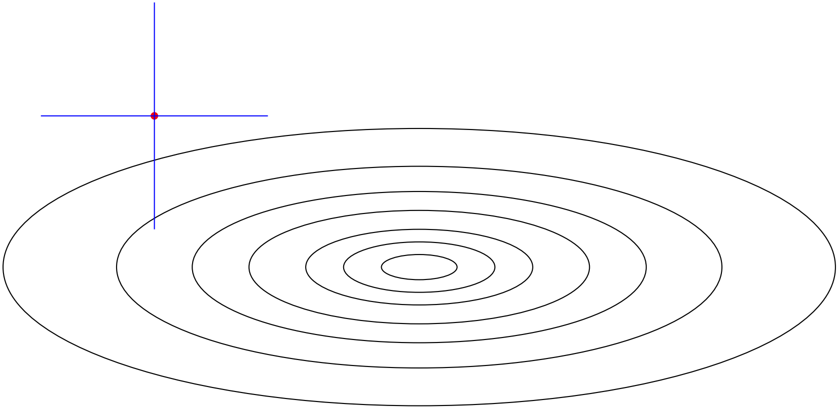
$$\begin{aligned} Z_m(\mathbf{z};\rho_k)) &= f^m(\mathbf{z}) - \rho_k^{\texttt{log}} \sum_{\ell \in \mathcal{G}^{\texttt{log}}} \log(-g_\ell^m(\mathbf{z})) + \\ &+ \frac{1}{\rho_k^{\texttt{ext}}} \left(\sum_{\ell \in \mathcal{G}^{\texttt{ext}}} (\max\{g_\ell^m(\mathbf{z}), 0\})^\nu + \sum_{j=1}^{p} |h_j^m(\mathbf{z})|^\nu \right), \end{aligned}$$

with $\nu \in (1, 2]$.

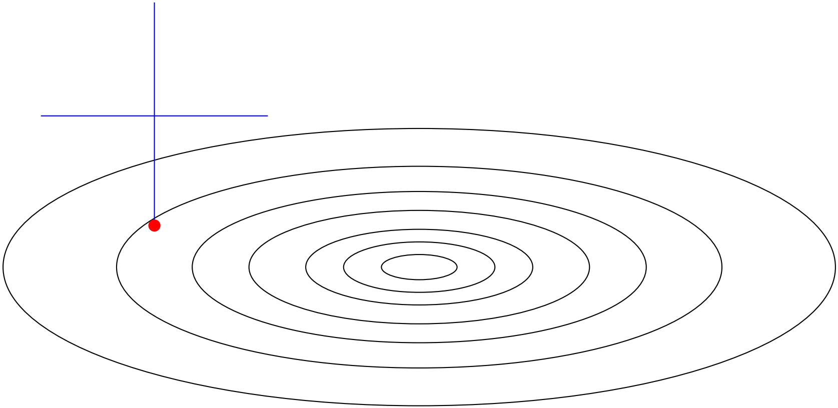
Poll Step in LOG-SID-PSM – Ordering



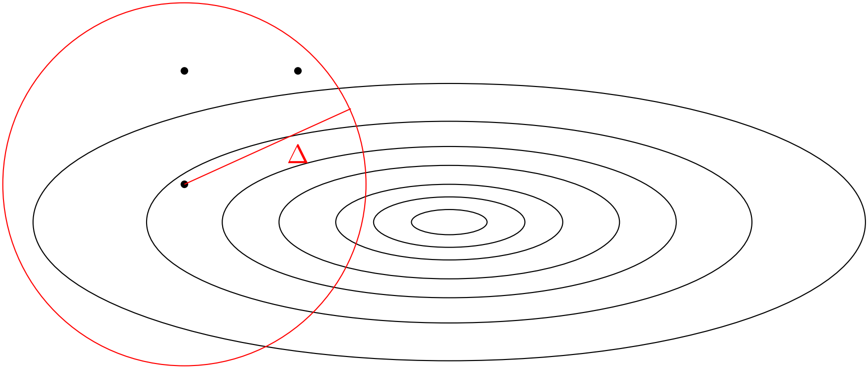
Poll Step in LOG-SID-PSM – Ordering



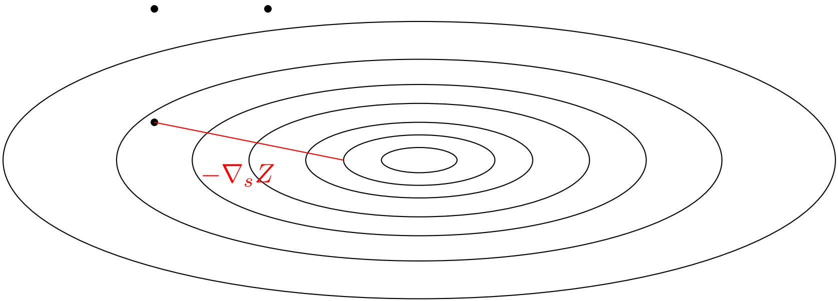
Poll Step in LOG-SID-PSM – Ordering



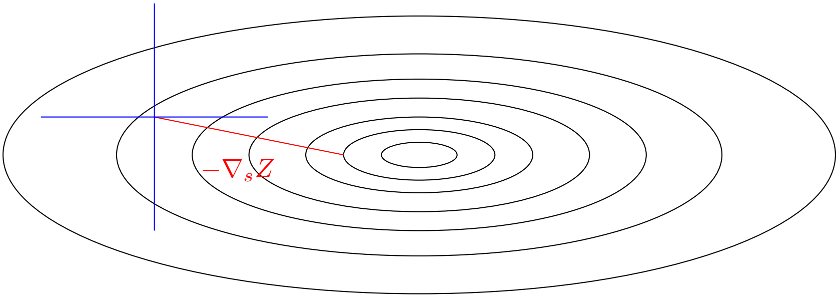
Poll Step in LOG-SID-PSM – Ordering



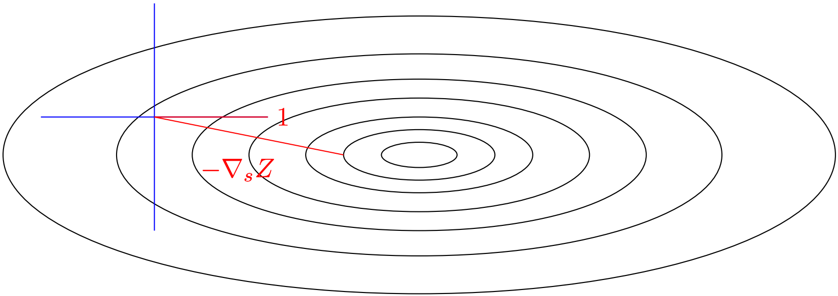
Poll Step in LOG-SID-PSM – Ordering



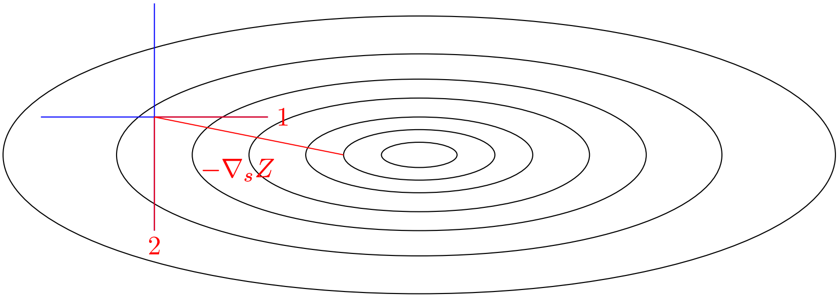
Poll Step in LOG-SID-PSM – Ordering



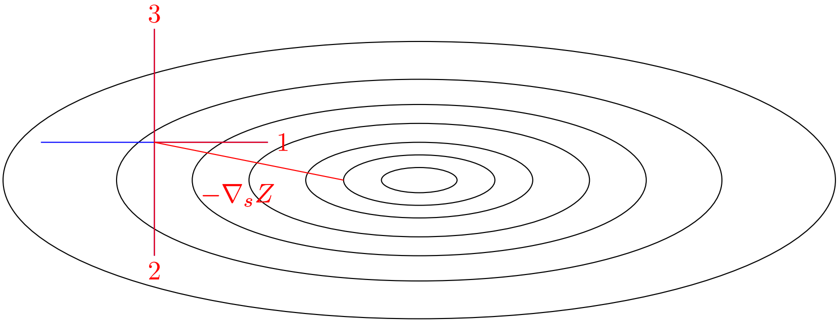
Poll Step in LOG-SID-PSM – Ordering



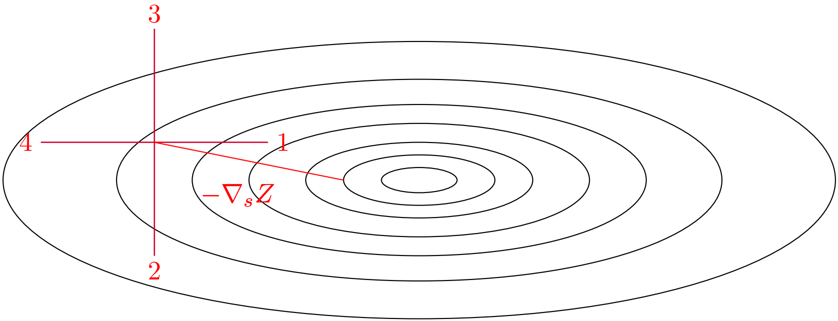
Poll Step in LOG-SID-PSM – Ordering



Poll Step in LOG-SID-PSM – Ordering



Poll Step in LOG-SID-PSM – Ordering



LOG-SID-PSM Penalty Parameters Rule

Penalty Function

$$Z(\mathbf{x};\rho_k) = f(\mathbf{x}) - \rho_k^{\texttt{log}} \sum_{\ell \in \mathcal{G}^{\texttt{log}}} \log(-g_\ell(\mathbf{x})) + \frac{1}{\rho_k^{\texttt{ext}}} \left(\sum_{\ell \in \mathcal{G}^{\texttt{ext}}} (\max\{0, g_\ell(\mathbf{x})\})^\nu + \sum_{j=1}^{p} |h_j(\mathbf{x})|^\nu \right)$$

- The updating rule splits into

Updating $\rho_k^{\texttt{log}}$

$$\alpha_{k+1} \leq \min\{(\rho_k^{\texttt{log}})^\beta, (g_{\min})_k^2\}$$

$$\Downarrow$$

$$\rho_{k+1}^{\texttt{log}} = \rho_k^{\texttt{log}} \min\{\eta, \max\{(g_{\min})_k^2, \zeta\}\}$$

Updating $\rho_k^{\texttt{ext}}$

$$\alpha_{k+1} \leq (\rho_k^{\texttt{ext}})^\beta$$

$$\Downarrow$$

$$\rho_{k+1}^{\texttt{ext}} = \min\left\{\zeta \rho_k^{\texttt{ext}}, \frac{\sqrt{\alpha_{k+1}}}{10}\right\}.$$

where $\beta = 1 + 10^{-9}$, $\nu = 1.1$, and $\zeta = 10^{-2}$.

Outline

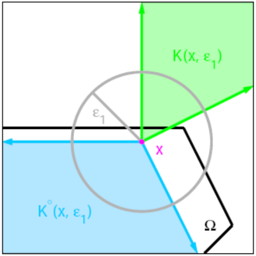
Numerical Settings

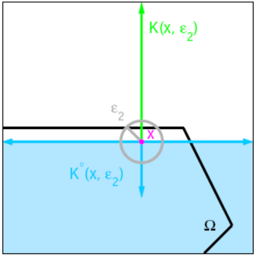
- 96 problems with nonlinear constraints and bounds from the CUTEst collection
  - number of variables between 1 and 50
  - number of nonlinear inequality constraints between 1 and 144
  - number of nonlinear equality constraints between 0 and 30
  - number of linear inequality constraints (other than bounds) between 0 and 123
- Initialization: provided in CUTEst collection
- Stopping criterion:
  - $\alpha_k < 10^{-8}$
  - Maximum of 2000 function evaluations

General Linear Constraints

Set of poll directions conforms to the geometry of nearby constraints.

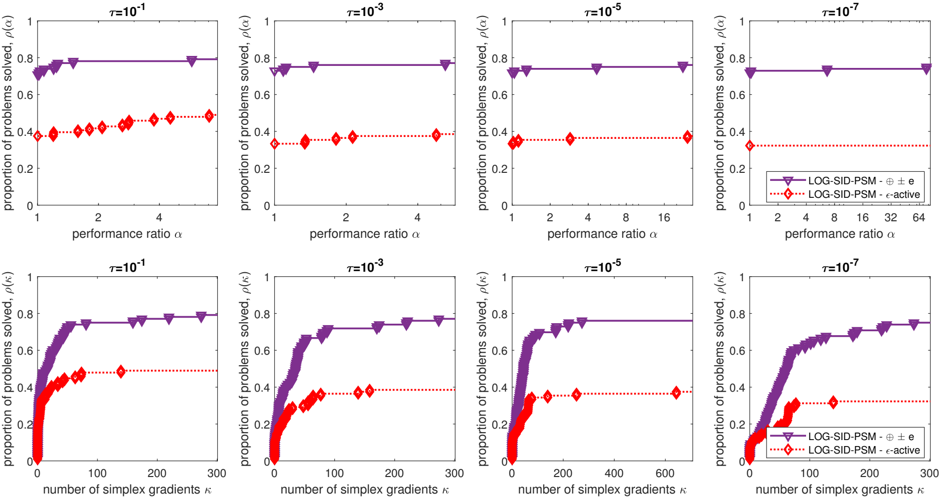
Implementation based in Abramson, Brezhneva, Dennis, and Pingel [2008].





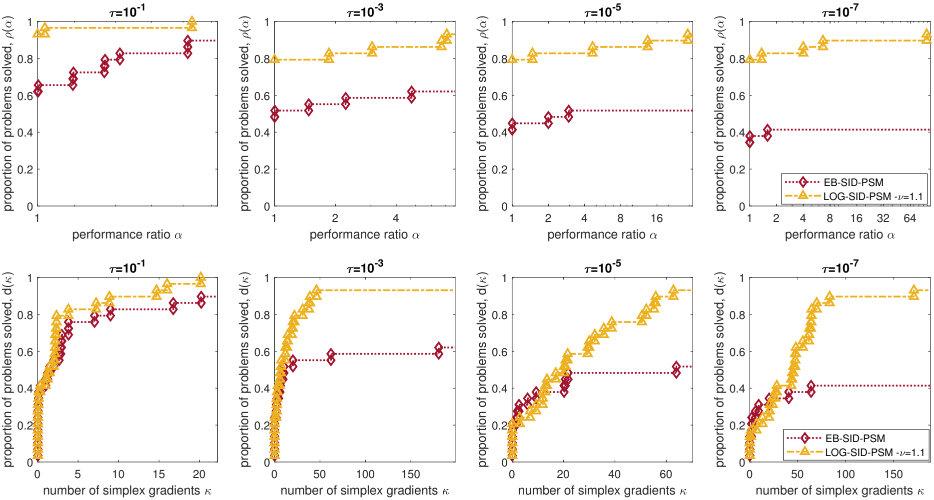
(in Kolda, Lewis, and Torczon [2003])

Comparison between strategies for linear constraints



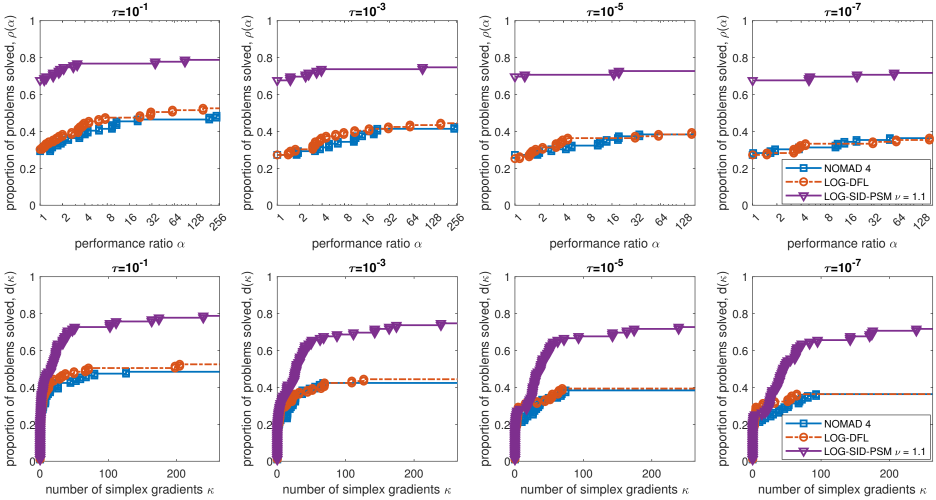
$^*\oplus \pm e \equiv [e \; -e \; I \; -I]$

LOG-SID-PSM vs SID-PSM

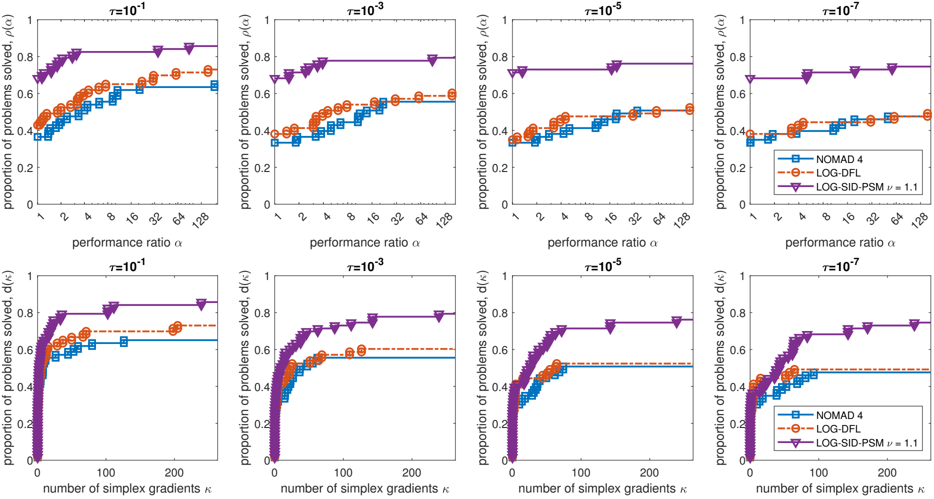


*28 problems from CUTEst

Comparison with other solvers (all problems)



Comparison with other solvers (only inequality constraints)



Outline

1. Introduction
2. Algorithmic Structure and Convergence Analysis
3. Implementation Details
4. Numerical Experiments
5. Conclusions and Future Work

Conclusions and Future Work

Conclusions

- LOG-SID-PSM keeps the basic algorithmic features of SID-PSM
- LOG-SID-PSM provides a mixed log-barrier approach to address nonlinear contraints in SID-PSM
- LOG-SID-PSM is competitive with state-of-the-art solvers for DFO problems with nonlinear constraints

Future Work

- Extension of the theoretical analysis for nonsmooth functions (*A. Brilli, A.L. Custódio, G. Liuzzi, and E.J. Silva*) **ongoing work**
- Extension of the constraint handling strategy to nonlinear derivative-free multiobjective optimization

Conclusions and Future Work

Conclusions

- LOG-SID-PSM keeps the basic algorithmic features of SID-PSM
- LOG-SID-PSM provides a mixed log-barrier approach to address nonlinear contraints in SID-PSM
- LOG-SID-PSM is competitive with state-of-the-art solvers for DFO problems with nonlinear constraints

Future Work

- Extension of the theoretical analysis for nonsmooth functions (*A. Brilli, A.L. Custódio, G. Liuzzi, and E.J. Silva*) **ongoing work**
- Extension of the constraint handling strategy to nonlinear derivative-free multiobjective optimization

Thank you for your attention!

Thechnical report will appear soon!