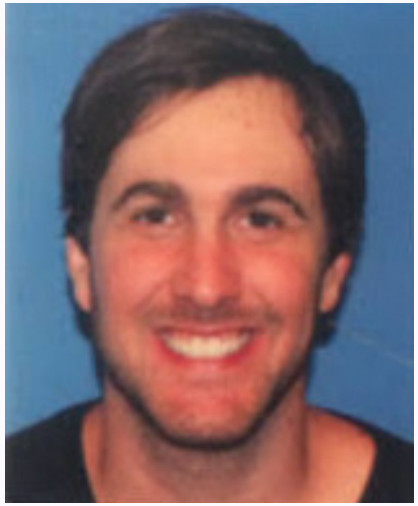
Positive spanning sets and resilient derivative-free optimization

Clément Royer

Modeling and Optimization: Theory and Applications

Ponta Delgada, June 17, 2025

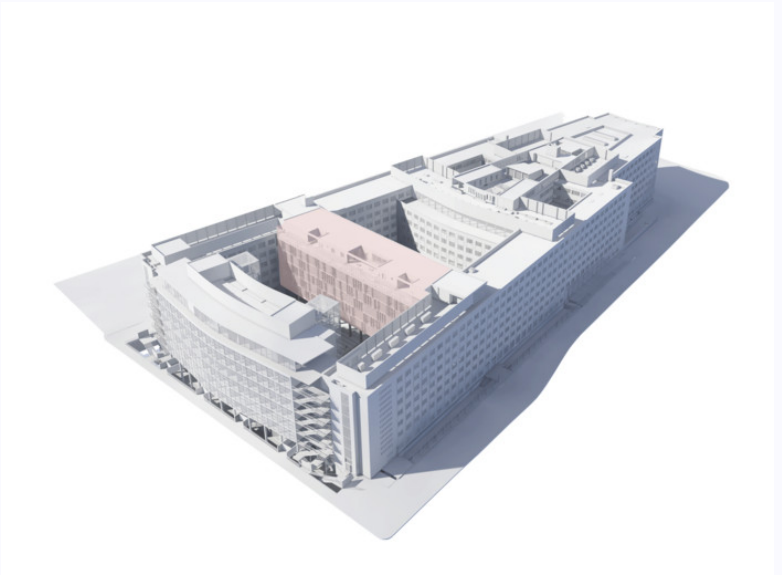
Collaborators


References

- W. Hare, G. Jarry-Bolduc, S. Kerleau and C. W. Royer.
  Using orthogonally structured positive bases for constructing positive k-spanning sets with cosine measure guarantees,
  Lin. Alg. Appl. 680:183-207, 2024.
- S. Kerleau and C. W. Royer.
  A derivative-free algorithm resilient to straggler function evaluations, Working paper.

Motivation: Dauphine's *Nouveau Campus*



- Building renovation, one wing at a time.
- Faculty/Staff to be moved during renovation.

Our task: Allocate office space during renovation.

Motivation: Dauphine's *Nouveau Campus* ('ed)

Our model for the Dauphine problem

- Huge integer LP, solved via Gurobi.
- $\sim$ 20 hyperparameters defining the model.
- Parallel runs on the department server.

Sub-task: Optimize hyperparameters to get the best LP optimal value.

Motivation: Dauphine's *Nouveau Campus* ('ed)

Our model for the Dauphine problem

- Huge integer LP, solved via Gurobi.
- $\sim$ 20 hyperparameters defining the model.
- Parallel runs on the department server.

Sub-task: Optimize hyperparameters to get the best LP optimal value.

Problem challenges

- Cannot differentiate (easily) within Gurobi.
  $\rightarrow$ **Derivative-free algorithms!**
- Cannot solve certain instances (48 hours for a feasible point!)
  $\rightarrow$ **Resiliency to straggler evaluations**.

Outline

Outline

Optimization problem

$$\text{minimize}_{x \in \mathbb{R}^n} f(x).$$

- f bounded below, nonconvex, $\mathcal{C}^{1,1}$ (for analysis).
- **Derivatives unavailable** for algorithmic use.
- Evaluations can take unusually long⇒**Stragglers**.

Optimization problem

$$\text{minimize}_{x \in \mathbb{R}^n} f(x).$$

- f bounded below, nonconvex, $\mathcal{C}^{1,1}$ (for analysis).
- **Derivatives unavailable** for algorithmic use.
- Evaluations can take unusually long⇒**Stragglers**.

Two families of DFO algorithms

- **Direct search**: Poll the space along directions.
- Model based: Build a model based on evaluations.

Direct search with stragglers

Inputs: $x_0 \in \mathbb{R}^n$, $\alpha_0 > 0$, $\mathcal{D} \subset \mathbb{R}^n$ set of nonzero vectors.
Iteration i: Given (x_i, α_i),

- If $\exists\ d_i \in \mathcal{D}$ such that

$$f(x_i + \alpha_i\, d_i) < f(x_i) - \alpha_i^2 \|d_i\|^2$$

  set $x_{i+1} := x_i + \alpha_i d_i$, $\alpha_{i+1} := 2\alpha_i$.
- Otherwise, set $x_{i+1} := x_i$, $\alpha_{i+1} := \alpha_i/2$.

Direct search with stragglers

Inputs: $x_0 \in \mathbb{R}^n$, $\alpha_0 > 0$, $\mathcal{D} \subset \mathbb{R}^n$ set of nonzero vectors.
Iteration i**:** Given (x_i, α_i),

- If $\exists\ d_i \in \mathcal{D} \setminus \mathcal{S}_i$ such that

$$f(x_i + \alpha_i\, d_i) < f(x_i) - \alpha_i^2 \|d_i\|^2$$

  set $x_{i+1} := x_i + \alpha_i d_i$, $\alpha_{i+1} := 2\alpha_i$.
- Otherwise, set $x_{i+1} := x_i$, $\alpha_{i+1} := \alpha_i/2$.

Straggler model: At every iteration i,

- $\exists \mathcal{S}_i \subset \mathcal{D}$ of **straggler directions** (f much longer to evaluate).
- $\mathcal{S}_i$ unknown before evaluations launched!
- Evaluations in $\mathcal{S}_i$ cannot be used in analyzing the method.

PSS and direct search

Positive spanning sets (PSS)

- For any $\mathcal{D} \subset \mathbb{R}^n$, the cosine measure of $\mathcal{D}$ is

$$\operatorname{cm}(\mathcal{D}) := \min_{v \neq 0} \max_{\substack{d \in \mathcal{D} \\ d \neq 0}} \frac{v^{\mathrm{T}} d}{\|v\| \|d\|}.$$

- $\mathcal{D}$ Positive Spanning Set $\iff \operatorname{cm}(\mathcal{D}) > 0$.

PSS and direct search

Positive spanning sets (PSS)

- For any $\mathcal{D} \subset \mathbb{R}^n$, the cosine measure of $\mathcal{D}$ is

$$\operatorname{cm}(\mathcal{D}) := \min_{v \neq 0} \max_{\substack{d \in \mathcal{D} \\ d \neq 0}} \frac{v^{\mathrm{T}} d}{\|v\| \|d\|}.$$

- $\mathcal{D}$ Positive Spanning Set $\iff \operatorname{cm}(\mathcal{D}) > 0$.

Guarantees for direct search with a PSS

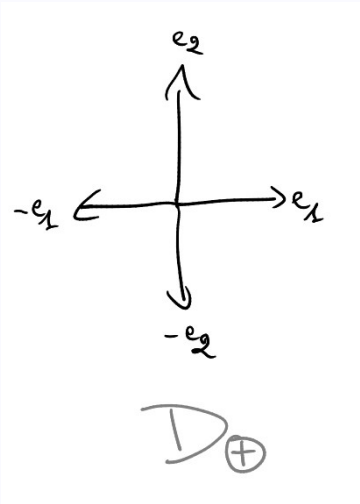
- Convergence: $\lim_i \|\nabla f(x_i)\| = 0$.
- Complexity: $\|\nabla f(x_l)\| \leq \epsilon$ after

$$\mathcal{O}\left(|\mathcal{D}| \operatorname{cm}(\mathcal{D})^{-2} \epsilon^{-2} \right)$$

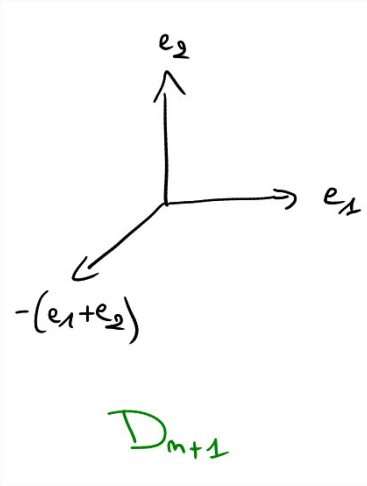
function evaluations.

Examples of positive spanning sets

$$\mathcal{D}_{\oplus} := \{e_1, \ldots, e_n, -e_1, \ldots, -e_n\}.$$
$$|\mathcal{D}_{\oplus}| = 2n, \ \mathrm{cm}(\mathcal{D}_{\oplus}) = \frac{1}{\sqrt{n}}.$$

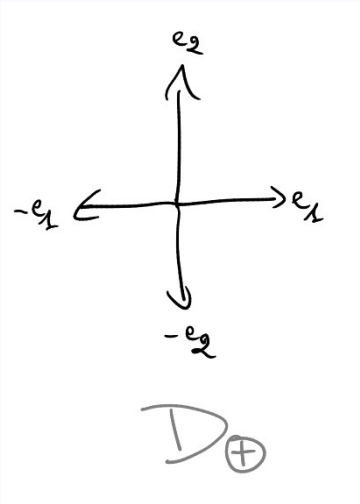


$$\mathcal{D}_{n+1} := \{e_1, \ldots, e_n, -\textstyle\sum_{l=1}^{n} e_l\}.$$
$$|\mathcal{D}_{n+1}| = n+1, \mathrm{cm}(\mathcal{D}_{n+1}) = \frac{1}{\sqrt{n^2+2(n-1)\sqrt{n}}}.$$

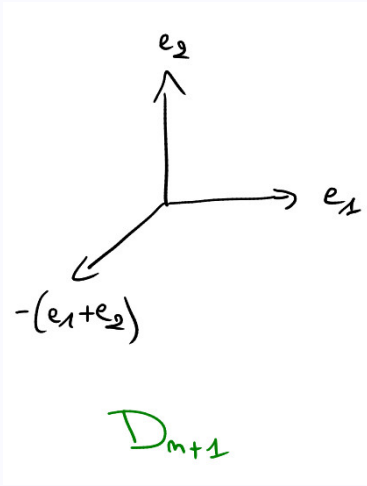


Examples of positive spanning sets

$$\mathcal{D}_{\oplus} := \{e_1, \ldots, e_n, -e_1, \ldots, -e_n\}.$$
$$|\mathcal{D}_{\oplus}| = 2n, \ \mathrm{cm}(\mathcal{D}_{\oplus}) = \frac{1}{\sqrt{n}}.$$



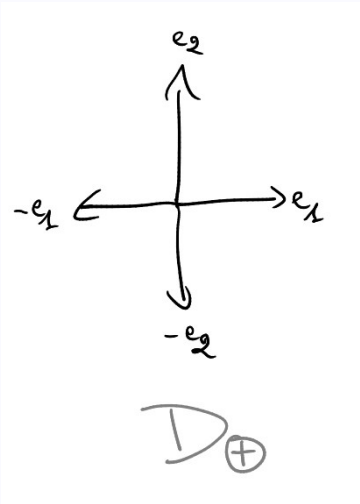
$$\mathcal{D}_{n+1} := \{e_1, \ldots, e_n, -\textstyle\sum_{l=1}^{n} e_l\}.$$
$$|\mathcal{D}_{n+1}| = n+1, \mathrm{cm}(\mathcal{D}_{n+1}) = \frac{1}{\sqrt{n^2+2(n-1)\sqrt{n}}}.$$

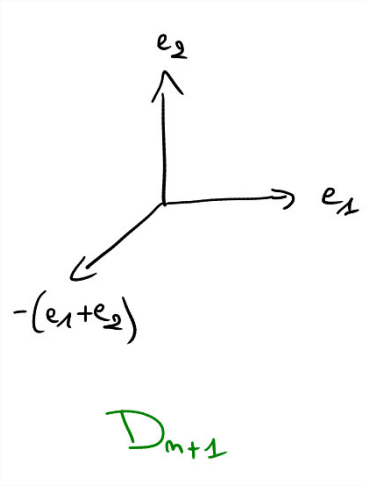


Examples of positive spanning sets

$\mathcal{D}_{\oplus} := \{e_1, \ldots, e_n, -e_1, \ldots, -e_n\}$.
$|\mathcal{D}_{\oplus}| = 2n$, $\mathrm{cm}(\mathcal{D}_{\oplus}) = \frac{1}{\sqrt{n}}$.

$\mathcal{D}_{n+1} := \{e_1, \ldots, e_n, -\sum_{l=1}^{n} e_l\}$.
$|\mathcal{D}_{n+1}| = n+1, \mathrm{cm}(\mathcal{D}_{n+1}) = \frac{1}{\sqrt{n^2+2(n-1)\sqrt{n}}}$.





Problem

- No proper subset of $\mathcal{D}_{\oplus}$ or $\mathcal{D}_{n+1}$ is a PSS!
- PSS: Not enough to guarantee convergence in presence of stragglers.

Outline

Positive k-spanning sets (PkSS)

Definition (Marcus '81-'84)

$\mathcal{D} \subset \mathbb{R}^n$ is a PkSS with $k \geq 1$ if

- Any $\mathcal{N} \subset \mathcal{D}$ with $|\mathcal{N}| = |\mathcal{D}| - k + 1$ is a PSS.
- Removing $k - 1$ vectors from $\mathcal{D}$ does not change its PSS nature.

Positive k-spanning sets (PkSS)

Definition (Marcus '81-'84)

$\mathcal{D} \subset \mathbb{R}^n$ is a PkSS with $k \geq 1$ if

- Any $\mathcal{N} \subset \mathcal{D}$ with $|\mathcal{N}| = |\mathcal{D}| - k + 1$ is a PSS.
- Removing $k-1$ vectors from $\mathcal{D}$ does not change its PSS nature.

The k-cosine measure (Hare, Jarry-Bolduc, Kerleau, R. '24)

- For any $\mathcal{D} \subset \mathbb{R}^n$, the k-cosine measure of $\mathcal{D}$ is

$$\operatorname{cm}_k(\mathcal{D}) = \min_{\substack{\mathcal{N} \subset \mathcal{D} \\ |\mathcal{N}| = |\mathcal{D}| - k + 1}} \operatorname{cm}(\mathcal{N}).$$

- $\mathcal{D}$ PkSS $\iff \operatorname{cm}_k(\mathcal{D}) > 0$.

Back to the algorithm

Inputs: $x_0 \in \mathbb{R}^n$, $\alpha_0 > 0$, $\mathcal{D} \subset \mathbb{R}^n$ **PkSS of nonzero vectors**.
Iteration i: Given (x_i, α_i),

- If $\exists\ d_i \in \mathcal{D} \setminus \mathcal{S}_i$ such that

$$f(x_i + \alpha_i\, d_i) < f(x_i) - \alpha_i^2 \|d_i\|^2$$

  set $x_{i+1} := x_i + \alpha_i d_i$, $\alpha_{i+1} := 2\alpha_i$.
- Otherwise, set $x_{i+1} := x_i$, $\alpha_{i+1} := \alpha_i/2$.

Back to the algorithm

Inputs: $x_0 \in \mathbb{R}^n$, $\alpha_0 > 0$, $\mathcal{D} \subset \mathbb{R}^n$ **PkSS of nonzero vectors**.
Iteration i: Given (x_i, α_i),

- If $\exists\ d_i \in \mathcal{D} \setminus \mathcal{S}_i$ such that

$$f(x_i + \alpha_i\, d_i) < f(x_i) - \alpha_i^2 \|d_i\|^2$$

  set $x_{i+1} := x_i + \alpha_i d_i$, $\alpha_{i+1} := 2\alpha_i$.
- Otherwise, set $x_{i+1} := x_i$, $\alpha_{i+1} := \alpha_i/2$.

Theorem (Kerleau, R. '25)

Assume $|\mathcal{S}_i| < k$ for all i. Then,

- $\liminf_i \|\nabla f(x_i)\| = 0$.
- $\|\nabla f(x_l)\| \leq \epsilon$ in at most $\mathcal{O}\left(|\mathcal{D}|\ \mathrm{cm}_k(\mathcal{D})^{-2}\epsilon^{-2}\right)$ function evaluations.

Outline

Building a PkSSfrom a PSS: First approach

- $\mathcal{D}_{\oplus} := \{e_1, \ldots, e_n, -e_1, \ldots, -e_n\}$, with $\{e_\ell\}$ coordinate basis vectors.
- $\beta\mathcal{D}_{\oplus}$: multiply all vectors by real β.

Building a PkSSfrom a PSS: First approach

- $\mathcal{D}_\oplus := \{e_1, \ldots, e_n, -e_1, \ldots, -e_n\}$, with $\{e_\ell\}$ coordinate basis vectors.
- $\beta\mathcal{D}_\oplus$: multiply all vectors by real β.

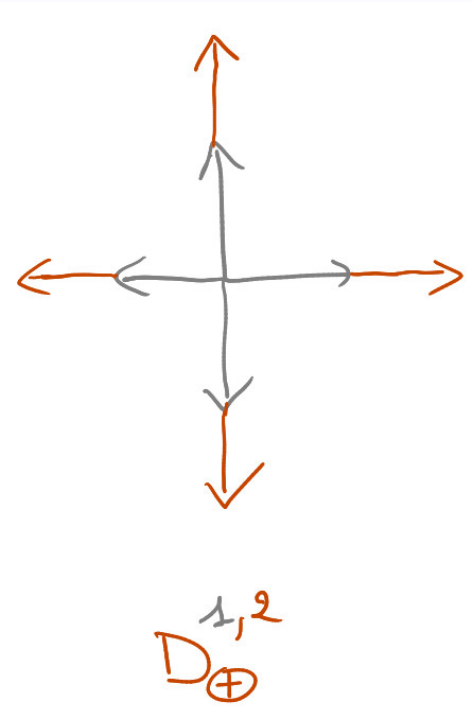
Duplicate vectors

If $\beta_1, \ldots, \beta_k > 0$ are distinct,

$$\mathcal{D}_\oplus^{\beta_{1:k}} := \bigcup_{j=1}^{k} \beta_j \mathcal{D}_\oplus$$

is a PkSS with

$$\mathrm{cm}_k(\mathcal{D}_\oplus^{\beta_{1:k}}) = \mathrm{cm}(\mathcal{D}_\oplus) = \frac{1}{\sqrt{n}}.$$



Building a PkSSfrom a PSS: Second approach

- $\mathcal{D}_{n+1} := \{e_1, \ldots, e_n, -\sum_{l=1}^{n} e_l\}$, with $\{e_\ell\}$ coordinate basis vectors.
- $R\mathcal{D}_{n+1}$: Apply **Rotation matrix** $R \in \mathbb{R}^{n \times n}$ to all vectors.

Building a PkSSfrom a PSS: Second approach

- $\mathcal{D}_{n+1} := \{e_1, \ldots, e_n, -\sum_{l=1}^{n} e_l\}$, with $\{e_\ell\}$ coordinate basis vectors.
- $R\mathcal{D}_{n+1}$: Apply **Rotation matrix** $R \in \mathbb{R}^{n\times n}$ to all vectors.

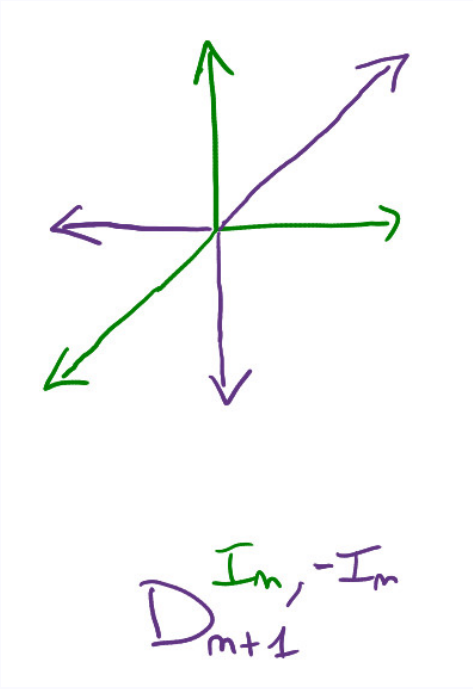
Rotate vectors

Let $R_1, \ldots, R_k$ be k distinct positive real numbers. The set

$$\mathcal{D}_{n+1}^{R_{1:k}} := \bigcup_{j=1}^{k} R_j \mathcal{D}_{n+1}$$

is a PkSS with

$$\mathrm{cm}_k(\mathcal{D}_{n+1}^{R_{1:k}}) \geq \mathrm{cm}(\mathcal{D}_{n+1}) = \frac{1}{\sqrt{n^2+2(n-1)\sqrt{n}}}.$$



Robust linear regression (Carmon et al '17)

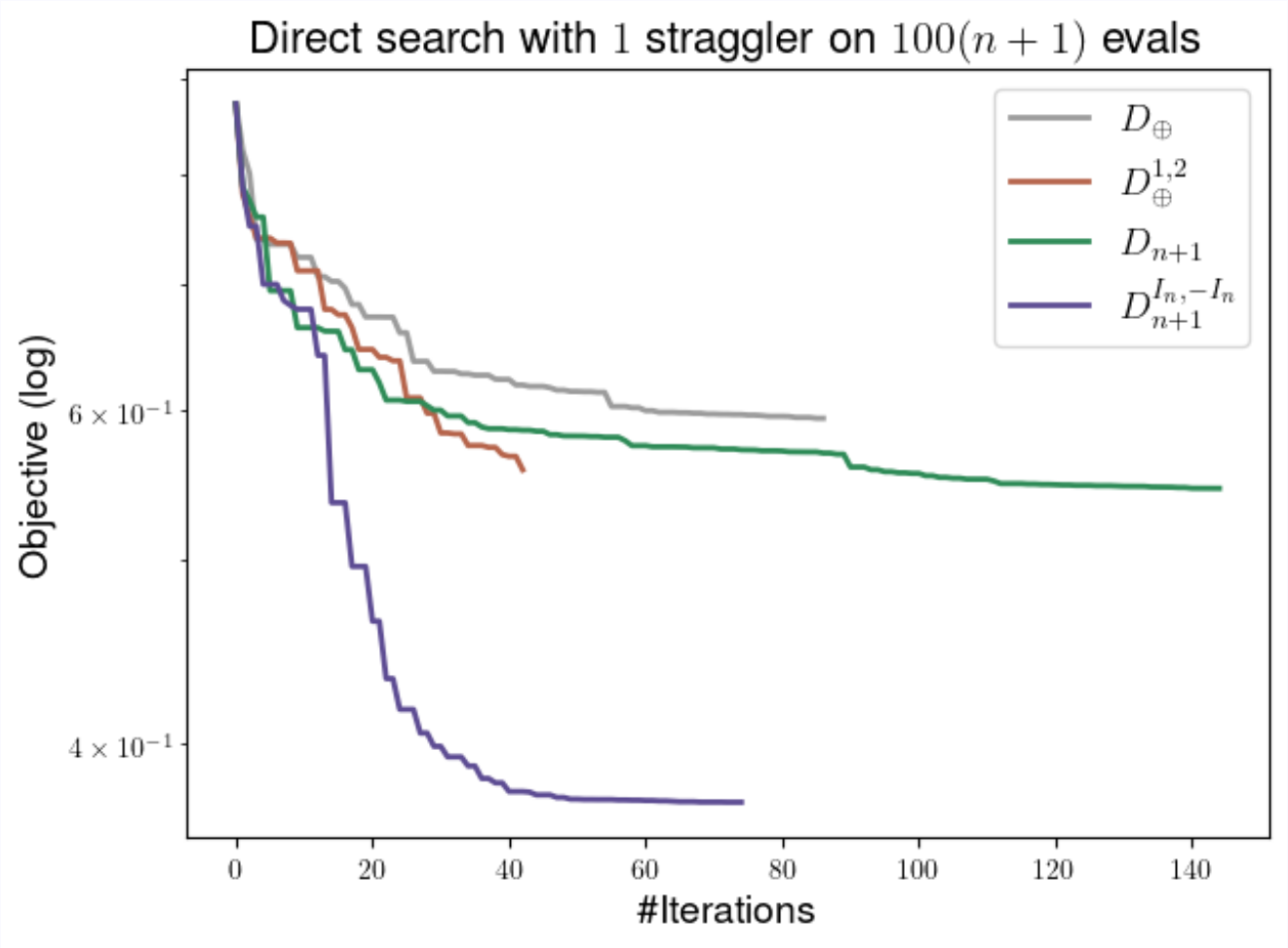
$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \frac{1}{2n}\sum_{i=1}^{2n}\phi(a_i^{\mathrm{T}}x-b_i),\quad \phi(t)=\frac{t^2}{1+t^2}.$$

- a_i i.i.d Gaussian, $b_i = a_i^{\mathrm{T}}z + 3u_1 + u_2$, $\{z, u_1\}$ Gaussian, u_2 Bernoulli.

Comparison: Direct search with PSS/P2SS, **one straggler/iteration**.

$\mathcal{D}$	$\lvert\mathcal{D}\rvert$	cm?
$\mathcal{D}_{\oplus}$	$2n$	$\text{cm} = \frac{1}{\sqrt{n}}$
$\mathcal{D}_{\oplus}^{1,2}$	$4n$	$\text{cm}_2 = \frac{1}{\sqrt{n}}$
$\mathcal{D}_{n+1}$	$n+1$	$\text{cm} = \frac{1}{\sqrt{n^2+2(n-1)\sqrt{n}}}$
$\mathcal{D}_{n+1}^{I_n,-I_n}$	$2n+2$	$\text{cm}_2 = \frac{1}{\sqrt{n^2+2(n-1)\sqrt{n}}}$.

Results in dimension 10



- P2SSs can outperform PSSs with stragglers!
- On 100 runs, $\mathcal{D}_{n+1}^{I_n,-I_n}$ gives best results.

Summing up

DFO with stragglers

- Resilient notion of PSS.
- Application to direct-search algorithms.

Moving forward

- More examples of PkSSs.
- Real experiments: Dauphine's problem.

Summing up

DFO with stragglers

- Resilient notion of PSS.
- Application to direct-search algorithms.

Moving forward

- More examples of PkSSs.
- Real experiments: Dauphine's problem.

Summing up

DFO with stragglers

- Resilient notion of PSS.
- Application to direct-search algorithms.

Moving forward

- More examples of PkSSs.
- Real experiments: Dauphine's problem.

Muito obrigado!

clement.royer@lamsade.dauphine.fr