

# Polling Set Construction and Worst-Case Complexity for Direct Search under Polyhedral Convex Constraints

*Joint work with Clément Royer (U. Paris Dauphine-PSL)*

---

Lindon Roberts, University of Melbourne ([lindon.roberts@unimelb.edu.au](mailto:lindon.roberts@unimelb.edu.au))

*Supported by the Australian Research Council (DE240100006) & CNRS IEA (BONUS)*

3rd Derivative-Free Optimization Symposium, Université de Montréal

25 June 2026

## Key references

This talk is based on:

- LR & C. W. Royer, Poll Set Construction and Worst-Case Complexity for Direct Search under Polyhedral Convex Constraints, arXiv:2605.27814 (2026).

Software available on Github: <https://github.com/lindonroberts/directsearch>

## Direct Search Methods

Considering **direct search methods** with sufficient decrease: [Kolda, Lewis & Torczon, 2003]

### Direct Search Iteration

- Given  $\mathbf{x}_k \in \mathbb{R}^n$  and  $\alpha_k > 0$ , choose a set  $\mathcal{D}_k \subset \mathbb{R}^n$  of  $p$  vectors
- If there exists  $\mathbf{d}_k \in \mathcal{D}_k$  with  $f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) < f(\mathbf{x}_k) - \frac{1}{2}\alpha_k^2$ 
  - Set  $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$  and increase  $\alpha_k$
  - Otherwise, set  $\mathbf{x}_{k+1} = \mathbf{x}_k$  and decrease  $\alpha_k$

# Direct Search Methods

Considering **direct search methods** with sufficient decrease: [Kolda, Lewis & Torczon, 2003]

## Direct Search Iteration

- Given  $\mathbf{x}_k \in \mathbb{R}^n$  and  $\alpha_k > 0$ , choose a set  $\mathcal{D}_k \subset \mathbb{R}^n$  of  $p$  vectors
- If there exists  $\mathbf{d}_k \in \mathcal{D}_k$  with  $f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) < f(\mathbf{x}_k) - \frac{1}{2}\alpha_k^2$ 
  - Set  $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$  and increase  $\alpha_k$
  - Otherwise, set  $\mathbf{x}_{k+1} = \mathbf{x}_k$  and decrease  $\alpha_k$

For convergence, need poll set  $\mathcal{D}_k$  to satisfy:

$$\max_{\mathbf{d} \in \mathcal{D}_k} \frac{\mathbf{d}^T (-\nabla f(\mathbf{x}_k))}{\|\mathbf{d}\|_2 \|\nabla f(\mathbf{x}_k)\|_2} \geq \kappa \in (0, 1]$$

i.e. there is a vector  $\mathbf{d}$  making a (sufficiently small) acute angle with  $-\nabla f(\mathbf{x}_k)$ .

## Direct Search: Convergence Theory

To quantify the quality of the poll set, look at the **cosine measure** of  $\mathcal{D}_k$ :

$$\text{cm}(\mathcal{D}_k) = \min_{\mathbf{v} \neq \mathbf{0}} \max_{\mathbf{d} \in \mathcal{D}_k} \frac{\mathbf{d}^T \mathbf{v}}{\|\mathbf{d}\|_2 \|\mathbf{v}\|_2}$$

i.e. (cosine of the) largest angle between  $\mathbf{v}$  and the closest (in angle)  $\mathbf{d} \in \mathcal{D}_k$  to  $\mathbf{v}$ .

## Direct Search: Convergence Theory

To quantify the quality of the poll set, look at the **cosine measure** of  $\mathcal{D}_k$ :

$$\text{cm}(\mathcal{D}_k) = \min_{\mathbf{v} \neq \mathbf{0}} \max_{\mathbf{d} \in \mathcal{D}_k} \frac{\mathbf{d}^T \mathbf{v}}{\|\mathbf{d}\|_2 \|\mathbf{v}\|_2}$$

i.e. (cosine of the) largest angle between  $\mathbf{v}$  and the closest (in angle)  $\mathbf{d} \in \mathcal{D}_k$  to  $\mathbf{v}$ .

### Theorem (Vicente, 2013)

*If  $f$  has Lipschitz continuous gradients,  $\|\mathbf{d}\|_2 \leq d_{\max}$  for all  $\mathbf{d} \in \mathcal{D}_k$  and  $\text{cm}(\mathcal{D}_k) \geq \kappa > 0$  for all  $k$ , then  $\|\nabla f(\mathbf{x}_k)\|_2 \leq \epsilon$  after at most  $\mathcal{O}(\kappa^{-2}\epsilon^{-2})$  iterations.*

## Direct Search: Convergence Theory

To quantify the quality of the poll set, look at the **cosine measure** of  $\mathcal{D}_k$ :

$$\text{cm}(\mathcal{D}_k) = \min_{\mathbf{v} \neq \mathbf{0}} \max_{\mathbf{d} \in \mathcal{D}_k} \frac{\mathbf{d}^T \mathbf{v}}{\|\mathbf{d}\|_2 \|\mathbf{v}\|_2}$$

i.e. (cosine of the) largest angle between  $\mathbf{v}$  and the closest (in angle)  $\mathbf{d} \in \mathcal{D}_k$  to  $\mathbf{v}$ .

### Theorem (Vicente, 2013)

*If  $f$  has Lipschitz continuous gradients,  $\|\mathbf{d}\|_2 \leq d_{\max}$  for all  $\mathbf{d} \in \mathcal{D}_k$  and  $\text{cm}(\mathcal{D}_k) \geq \kappa > 0$  for all  $k$ , then  $\|\nabla f(\mathbf{x}_k)\|_2 \leq \epsilon$  after at most  $\mathcal{O}(\kappa^{-2}\epsilon^{-2})$  iterations.*

For  $\pm$  coordinate directions,  $\kappa = 1/\sqrt{n} \rightarrow \mathcal{O}(n\epsilon^{-2})$  iterations, or  $\mathcal{O}(n^2\epsilon^{-2})$  objective evaluations (up to  $|\mathcal{D}_k| = 2n$  evaluations per iteration).

## An alternative perspective

**Link 1:** A classical result is:

### Theorem (Davis, 1954)

*A set  $\mathcal{D}$  has  $\text{cm}(\mathcal{D}) > 0$  if and only if  $\mathcal{D}$  is a **positive spanning set (PSS)** (i.e. can write every  $\mathbf{v} \in \mathbb{R}^n$  as a positive combination of vectors in  $\mathcal{D}$ , or  $\text{cone}(\mathcal{D}) = \mathbb{R}^n$ ).*

# An alternative perspective

**Link 1:** A classical result is:

## Theorem (Davis, 1954)

A set  $\mathcal{D}$  has  $\text{cm}(\mathcal{D}) > 0$  if and only if  $\mathcal{D}$  is a *positive spanning set (PSS)* (i.e. can write every  $\mathbf{v} \in \mathbb{R}^n$  as a positive combination of vectors in  $\mathcal{D}$ , or  $\text{cone}(\mathcal{D}) = \mathbb{R}^n$ ).

**Link 2:** In **model-based DFO**, suppose we build a linear interpolant  $m(\cdot)$  using perturbations  $\mathbf{d}_1, \dots, \mathbf{d}_n$  around  $\mathbf{x}$ . Then

$$|m(\mathbf{x} + \mathbf{d}) - f(\mathbf{x}) - \nabla f(\mathbf{x})^T \mathbf{d}| \leq \frac{L_{\nabla f}}{2} d_{\max}^2 \sum_{i=1}^n |c_i(\mathbf{d})|,$$

where the  $c_i(\mathbf{d})$  satisfy  $\mathbf{d} = \sum_{i=1}^n c_i(\mathbf{d}) \mathbf{d}_i$  (exist whenever  $\mathbf{d}_i$  linearly span  $\mathbb{R}^n$ ).

This corresponds to  **$\Lambda$ -poisedness** or **Lebesgue measure**.

[Trefethen, 2020], [LR, 2025]

## Positive Spanning Sets

These two ideas motivate an alternative measure of poll set quality, generalizing the Lebesgue measure to positive spanning sets.

### Definition

A set  $\mathcal{D} = \{\mathbf{d}_1, \dots, \mathbf{d}_p\}$  is a  $\Lambda$ -positive spanning set ( $\Lambda$ -PSS) for  $B(\mathbf{x}, \alpha)$  if, for any  $\|\mathbf{v}\|_2 \leq \alpha$ , we can write  $\mathbf{v} = \sum_{i=1}^p c_i(\mathbf{v})\mathbf{d}_i$  with  $c_i(\mathbf{v}) \geq 0$  and  $\sum_{i=1}^p c_i(\mathbf{v}) \leq \Lambda$ .

## Positive Spanning Sets

These two ideas motivate an alternative measure of poll set quality, generalizing the Lebesgue measure to positive spanning sets.

### Definition

A set  $\mathcal{D} = \{\mathbf{d}_1, \dots, \mathbf{d}_p\}$  is a  $\Lambda$ -positive spanning set ( $\Lambda$ -PSS) for  $B(\mathbf{x}, \alpha)$  if, for any  $\|\mathbf{v}\|_2 \leq \alpha$ , we can write  $\mathbf{v} = \sum_{i=1}^p c_i(\mathbf{v})\mathbf{d}_i$  with  $c_i(\mathbf{v}) \geq 0$  and  $\sum_{i=1}^p c_i(\mathbf{v}) \leq \Lambda$ .

This is essentially equivalent to a lower bound  $\text{cm}(\mathcal{D}) \geq \kappa$  with  $\kappa = 1/\Lambda$ .

### Theorem (LR & Royer, 2026)

- (a) If  $\mathcal{D}$  is a  $\Lambda$ -PSS and  $\|\mathbf{d}_i\|_2 \leq d_{\max}\alpha$ , then  $\text{cm}(\mathcal{D}) \geq 1/(d_{\max}\Lambda)$
- (b) If  $\text{cm}(\mathcal{D}) \geq \kappa$  and  $\|\mathbf{d}_i\|_2 \geq d_{\min}\alpha$ , then  $\mathcal{D}$  is a  $\Lambda$ -PSS with  $\Lambda = 1/(d_{\min}\kappa)$ .

The earlier result (Davis, 1954) corresponds to the limit  $\kappa \rightarrow 0^+$  or  $\Lambda \rightarrow \infty$ .

## Convex Constraints

Now, consider the constrained problem

$$\min_{\mathbf{x} \in \Omega} f(\mathbf{x}),$$

where  $\Omega \subseteq \mathbb{R}^n$  is a convex set with nonempty interior (e.g. bounds, linear inequalities).

## Convex Constraints

Now, consider the constrained problem

$$\min_{\mathbf{x} \in \Omega} f(\mathbf{x}),$$

where  $\Omega \subseteq \mathbb{R}^n$  is a convex set with nonempty interior (e.g. bounds, linear inequalities).

One measure of first-order optimality for this problem is

$$\pi(\mathbf{x}) := \left| \min_{\substack{\mathbf{x} + \mathbf{v} \in \Omega \\ \|\mathbf{v}\|_2 \leq 1}} \nabla f(\mathbf{x})^T \mathbf{v} \right| \quad (\text{e.g. } \pi(\mathbf{x}) = \|\nabla f(\mathbf{x})\|_2 \text{ if } \Omega = \mathbb{R}^n)$$

We have:  $\pi(\mathbf{x}) \geq 0$ , with  $\pi(\mathbf{x}) = 0$  if and only if  $\mathbf{x}$  is first-order critical, and  $\pi(\cdot)$  is continuous.

e.g. [Cartis, Gould & Toint, 2012]

## Convex Constraints

Now, consider the constrained problem

$$\min_{\mathbf{x} \in \Omega} f(\mathbf{x}),$$

where  $\Omega \subseteq \mathbb{R}^n$  is a convex set with nonempty interior (e.g. bounds, linear inequalities).

One measure of first-order optimality for this problem is

$$\pi(\mathbf{x}) := \left| \min_{\substack{\mathbf{x} + \mathbf{v} \in \Omega \\ \|\mathbf{v}\|_2 \leq 1}} \nabla f(\mathbf{x})^T \mathbf{v} \right| \quad (\text{e.g. } \pi(\mathbf{x}) = \|\nabla f(\mathbf{x})\|_2 \text{ if } \Omega = \mathbb{R}^n)$$

We have:  $\pi(\mathbf{x}) \geq 0$ , with  $\pi(\mathbf{x}) = 0$  if and only if  $\mathbf{x}$  is first-order critical, and  $\pi(\cdot)$  is continuous. e.g. [Cartis, Gould & Toint, 2012]

**Question:** how to generalize  $\text{cm}(\mathcal{D})$  to the constrained setting?

## Constrained Positive Spanning Sets

There is no natural generalization of  $\text{cm}(\mathcal{D})$  to the constrained setting\*, but there is for  $\Lambda$ -PSS.

### Definition

A set  $\mathcal{D} = \{\mathbf{d}_1, \dots, \mathbf{d}_p\}$  is a  $\Lambda$ -positive spanning set ( $\Lambda$ -PSS) for  $B(\mathbf{x}, \alpha) \cap \Omega$  if, for any  $\|\mathbf{v}\|_2 \leq \alpha$  with  $\mathbf{x} + \mathbf{v} \in \Omega$ , we can write  $\mathbf{v} = \sum_{i=1}^p c_i(\mathbf{v}) \mathbf{d}_i$  with  $c_i(\mathbf{v}) \geq 0$  and  $\sum_{i=1}^p c_i(\mathbf{v}) \leq \Lambda$ .

i.e. just ensure all relevant points are feasible.

## Constrained Positive Spanning Sets

There is no natural generalization of  $\text{cm}(\mathcal{D})$  to the constrained setting\*, but there is for  $\Lambda$ -PSS.

### Definition

A set  $\mathcal{D} = \{\mathbf{d}_1, \dots, \mathbf{d}_p\}$  is a  $\Lambda$ -positive spanning set ( $\Lambda$ -PSS) for  $B(\mathbf{x}, \alpha) \cap \Omega$  if, for any  $\|\mathbf{v}\|_2 \leq \alpha$  with  $\mathbf{x} + \mathbf{v} \in \Omega$ , we can write  $\mathbf{v} = \sum_{i=1}^p c_i(\mathbf{v}) \mathbf{d}_i$  with  $c_i(\mathbf{v}) \geq 0$  and  $\sum_{i=1}^p c_i(\mathbf{v}) \leq \Lambda$ .

i.e. just ensure all relevant points are feasible.

\*only for subsets of the constrained region (approximate tangent cone — discussed later).

# Constrained Positive Spanning Sets

This enables a very simple worst-case complexity result.

## Theorem (LR & Royer, 2026)

*If  $f$  has Lipschitz continuous gradients,  $\|\mathbf{d}\|_2 \leq d_{\max}\alpha_k$  for all  $\mathbf{d} \in \mathcal{D}_k$  and  $\mathcal{D}_k$  is a  $\Lambda$ -PSS for  $B(\mathbf{x}_k, \alpha_k) \cap \Omega$  for all  $k$ , then  $\pi(\mathbf{x}_k) \leq \epsilon$  after at most  $\mathcal{O}(\Lambda^2\epsilon^{-2})$  iterations.*

# Constrained Positive Spanning Sets

This enables a very simple worst-case complexity result.

## Theorem (LR & Royer, 2026)

*If  $f$  has Lipschitz continuous gradients,  $\|\mathbf{d}\|_2 \leq d_{\max}\alpha_k$  for all  $\mathbf{d} \in \mathcal{D}_k$  and  $\mathcal{D}_k$  is a  $\Lambda$ -PSS for  $B(\mathbf{x}_k, \alpha_k) \cap \Omega$  for all  $k$ , then  $\pi(\mathbf{x}_k) \leq \epsilon$  after at most  $\mathcal{O}(\Lambda^2\epsilon^{-2})$  iterations.*

Key new idea: by definition,  $\pi(\mathbf{x}_k) = -\nabla f(\mathbf{x}_k)^T \mathbf{v}_k$  for some  $\|\mathbf{v}_k\|_2 \leq 1$  with  $\mathbf{x}_k + \mathbf{v}_k \in \Omega$ , invoke  $\Lambda$ -PSS property (or  $\alpha_k \mathbf{v}_k$  if  $\alpha_k < 1$ ).

# Constrained Positive Spanning Sets

This enables a very simple worst-case complexity result.

## Theorem (LR & Royer, 2026)

*If  $f$  has Lipschitz continuous gradients,  $\|\mathbf{d}\|_2 \leq d_{\max}\alpha_k$  for all  $\mathbf{d} \in \mathcal{D}_k$  and  $\mathcal{D}_k$  is a  $\Lambda$ -PSS for  $B(\mathbf{x}_k, \alpha_k) \cap \Omega$  for all  $k$ , then  $\pi(\mathbf{x}_k) \leq \epsilon$  after at most  $\mathcal{O}(\Lambda^2\epsilon^{-2})$  iterations.*

Key new idea: by definition,  $\pi(\mathbf{x}_k) = -\nabla f(\mathbf{x}_k)^T \mathbf{v}_k$  for some  $\|\mathbf{v}_k\|_2 \leq 1$  with  $\mathbf{x}_k + \mathbf{v}_k \in \Omega$ , invoke  $\Lambda$ -PSS property (or  $\alpha_k \mathbf{v}_k$  if  $\alpha_k < 1$ ).

This matches the unconstrained result with  $\Lambda \sim 1/\kappa$ , and derivative-based and interpolation-based DFO theory. [Cartis, Gould & Toint, 2012], [Hough & LR, 2022]

# Bound Constraints

## Problem

How do we construct  $\Lambda$ -PSS sets for constrained problems?

# Bound Constraints

## Problem

How do we construct  $\Lambda$ -PSS sets for constrained problems?

This is easy for bound-constrained problems.

# Bound Constraints

## Problem

How do we construct  $\Lambda$ -PSS sets for constrained problems?

This is easy for bound-constrained problems.

If we have  $\Omega = \{\mathbf{x} : \mathbf{x}^L \leq \mathbf{x} \leq \mathbf{x}^U\}$ , then use  $\pm$  coordinate directions, scaled to length  $\alpha$  and feasibility:

$$\mathcal{D} = \cup_{i=1}^n \{\alpha_i \mathbf{e}_i, -\alpha_{-i} \mathbf{e}_i\}, \quad \text{where, e.g. } \alpha_i = \min(\alpha, x_i^U - x_i)$$

# Bound Constraints

## Problem

How do we construct  $\Lambda$ -PSS sets for constrained problems?

This is easy for bound-constrained problems.

If we have  $\Omega = \{\mathbf{x} : \mathbf{x}^L \leq \mathbf{x} \leq \mathbf{x}^U\}$ , then use  $\pm$  coordinate directions, scaled to length  $\alpha$  and feasibility:

$$\mathcal{D} = \cup_{i=1}^n \{\alpha_i \mathbf{e}_i, -\alpha_{-i} \mathbf{e}_i\}, \quad \text{where, e.g. } \alpha_i = \min(\alpha, x_i^U - x_i)$$

## Theorem (LR & Royer, 2026)

$\mathcal{D}$  is a  $\Lambda$ -PSS with  $\sqrt{n} \leq \Lambda(\mathbf{x}, \alpha) \leq n$  (depending on # nearly active constraints).

# Bound Constraints

## Problem

How do we construct  $\Lambda$ -PSS sets for constrained problems?

This is easy for bound-constrained problems.

If we have  $\Omega = \{\mathbf{x} : \mathbf{x}^L \leq \mathbf{x} \leq \mathbf{x}^U\}$ , then use  $\pm$  coordinate directions, scaled to length  $\alpha$  and feasibility:

$$\mathcal{D} = \cup_{i=1}^n \{\alpha_i \mathbf{e}_i, -\alpha_{-i} \mathbf{e}_i\}, \quad \text{where, e.g. } \alpha_i = \min(\alpha, x_i^U - x_i)$$

## Theorem (LR & Royer, 2026)

$\mathcal{D}$  is a  $\Lambda$ -PSS with  $\sqrt{n} \leq \Lambda(\mathbf{x}, \alpha) \leq n$  (depending on # nearly active constraints).

This leads to  $\mathcal{O}(n^2 \epsilon^{-2})$  iterations or  $\mathcal{O}(n^3 \epsilon^{-2})$  evaluations to achieve  $\pi(\mathbf{x}_k) \leq \epsilon$  [note:  $\mathcal{O}(n^3)$  is not optimal result, but better in practice]. [Gratton et al., 2019]

## Linearly Constrained Problems

Now suppose our feasible set is a polytope, formed by  $\mathbf{a}_j^T \mathbf{x} \leq b_j$  for  $j = 1, \dots, m$ .

**How is linearly constrained direct search currently performed?**

# Linearly Constrained Problems

Now suppose our feasible set is a polytope, formed by  $\mathbf{a}_j^T \mathbf{x} \leq b_j$  for  $j = 1, \dots, m$ .

**How is linearly constrained direct search currently performed?**

At every iteration (with  $\mathbf{x}_k$  feasible), look at the **nearly active constraints**:

$$j \in \mathcal{J}(\mathbf{x}_k, \alpha_k) \quad \iff \quad b_j - \alpha_k \|\mathbf{a}_j\|_2^2 \leq \mathbf{a}_j^T \mathbf{x}_k \leq b_j$$

i.e. the constraints whose boundaries intersect with  $B(\mathbf{x}_k, \alpha_k)$ .

# Linearly Constrained Problems

Now suppose our feasible set is a polytope, formed by  $\mathbf{a}_j^T \mathbf{x} \leq b_j$  for  $j = 1, \dots, m$ .

**How is linearly constrained direct search currently performed?**

At every iteration (with  $\mathbf{x}_k$  feasible), look at the **nearly active constraints**:

$$j \in \mathcal{J}(\mathbf{x}_k, \alpha_k) \quad \iff \quad b_j - \alpha_k \|\mathbf{a}_j\|_2^2 \leq \mathbf{a}_j^T \mathbf{x}_k \leq b_j$$

i.e. the constraints whose boundaries intersect with  $B(\mathbf{x}_k, \alpha_k)$ .

The **approximate normal cone**  $N_\Omega(\mathbf{x}_k, \alpha_k)$  is generated by  $\mathbf{a}_j$  for  $j \in \mathcal{J}(\mathbf{x}_k, \alpha_k)$ .

# Linearly Constrained Problems

Now suppose our feasible set is a polytope, formed by  $\mathbf{a}_j^T \mathbf{x} \leq b_j$  for  $j = 1, \dots, m$ .

**How is linearly constrained direct search currently performed?**

At every iteration (with  $\mathbf{x}_k$  feasible), look at the **nearly active constraints**:

$$j \in \mathcal{J}(\mathbf{x}_k, \alpha_k) \quad \iff \quad b_j - \alpha_k \|\mathbf{a}_j\|_2^2 \leq \mathbf{a}_j^T \mathbf{x}_k \leq b_j$$

i.e. the constraints whose boundaries intersect with  $B(\mathbf{x}_k, \alpha_k)$ .

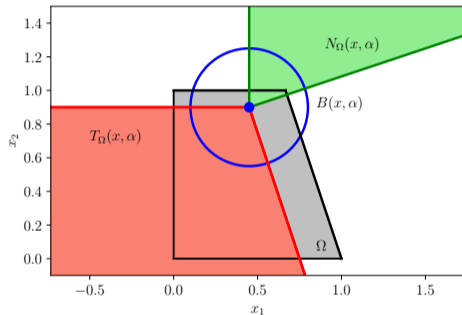
The **approximate normal cone**  $N_\Omega(\mathbf{x}_k, \alpha_k)$  is generated by  $\mathbf{a}_j$  for  $j \in \mathcal{J}(\mathbf{x}_k, \alpha_k)$ .

The **approximate tangent cone** is  $T_\Omega(\mathbf{x}_k, \alpha_k)$  is its polar:  $\mathbf{v}$  such that  $\mathbf{v}^T \mathbf{a}_j \leq 0$  for all  $j \in \mathcal{J}(\mathbf{x}_k, \alpha_k)$ .

[Kolda, Lewis & Torczon, 2003 & 2007]

## Example: Direct Search

Simple example:  $x_1, x_2 \geq 0$ , with  $x_2 \leq 1$  and  $3x_1 + x_2 \leq 3$  at  $\mathbf{x} = [0.45, 0.9]$ :

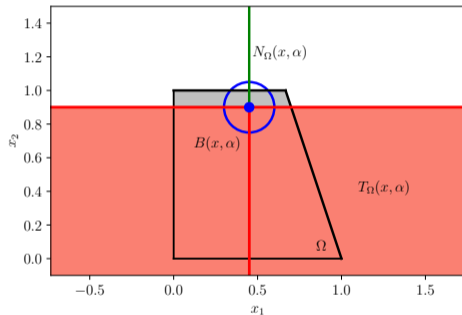


(a)  $\alpha = 0.35$ , nearly active constraints are  $x_2 \leq 1$  and  $3x_1 + x_2 \leq 3$

Modified from [Kolda, Lewis & Torczon, 2007]

## Example: Direct Search

Simple example:  $x_1, x_2 \geq 0$ , with  $x_2 \leq 1$  and  $3x_1 + x_2 \leq 3$  at  $\mathbf{x} = [0.45, 0.9]$ :



**(b)  $\alpha = 0.15$ , nearly active constraint is  $x_2 \leq 1$**

Modified from [Kolda, Lewis & Torczon, 2007]

# Linearly Constrained Problems

In existing theory, we choose the poll set such that

$$\mathcal{D}_k \supseteq \text{generators of } T_{\Omega}(\mathbf{x}_k, \alpha_k)$$

such that

$$\text{cm}_{\mathcal{T}}(\mathcal{D}) = \min_{\text{proj}_{T_{\Omega}(\mathbf{x}_k, \alpha_k)}(\mathbf{v}) \neq \mathbf{0}} \max_{\mathbf{d} \in \mathcal{D}} \frac{\mathbf{d}^T \mathbf{v}}{\|\mathbf{d}\|_2 \|\text{proj}_{T_{\Omega}(\mathbf{x}_k, \alpha_k)}(\mathbf{v})\|_2} > 0$$

# Linearly Constrained Problems

In existing theory, we choose the poll set such that

$$\mathcal{D}_k \supseteq \text{generators of } T_{\Omega}(\mathbf{x}_k, \alpha_k)$$

such that

$$\text{cm}_{\mathcal{T}}(\mathcal{D}) = \min_{\text{proj}_{T_{\Omega}(\mathbf{x}_k, \alpha_k)}(\mathbf{v}) \neq \mathbf{0}} \max_{\mathbf{d} \in \mathcal{D}} \frac{\mathbf{d}^T \mathbf{v}}{\|\mathbf{d}\|_2 \|\text{proj}_{T_{\Omega}(\mathbf{x}_k, \alpha_k)}(\mathbf{v})\|_2} > 0$$

## Theorem (Gratton et al., 2019)

*If  $f$  has Lipschitz continuous gradients,  $\|\mathbf{d}\|_2 \leq d_{\max}$  for all  $\mathbf{d} \in \mathcal{D}_k$  and  $\text{cm}_{\mathcal{T}}(\mathcal{D}_k) \geq \kappa > 0$  for all  $k$ , then  $\pi(\mathbf{x}_k) \leq \epsilon$  after at most  $\mathcal{O}(\kappa^{-2}\epsilon^{-2})$  iterations.*

# Linearly Constrained Problems

In existing theory, we choose the poll set such that

$$\mathcal{D}_k \supseteq \text{generators of } T_{\Omega}(\mathbf{x}_k, \alpha_k)$$

such that

$$\text{cm}_{\mathcal{T}}(\mathcal{D}) = \min_{\text{proj}_{T_{\Omega}(\mathbf{x}_k, \alpha_k)}(\mathbf{v}) \neq \mathbf{0}} \max_{\mathbf{d} \in \mathcal{D}} \frac{\mathbf{d}^T \mathbf{v}}{\|\mathbf{d}\|_2 \|\text{proj}_{T_{\Omega}(\mathbf{x}_k, \alpha_k)}(\mathbf{v})\|_2} > 0$$

## Theorem (Gratton et al., 2019)

*If  $f$  has Lipschitz continuous gradients,  $\|\mathbf{d}\|_2 \leq d_{\max}$  for all  $\mathbf{d} \in \mathcal{D}_k$  and  $\text{cm}_{\mathcal{T}}(\mathcal{D}_k) \geq \kappa > 0$  for all  $k$ , then  $\pi(\mathbf{x}_k) \leq \epsilon$  after at most  $\mathcal{O}(\kappa^{-2}\epsilon^{-2})$  iterations.*

For bound constraints, we can get  $\kappa = 1/\sqrt{n}$ , matching the unconstrained case.

### Why bother?

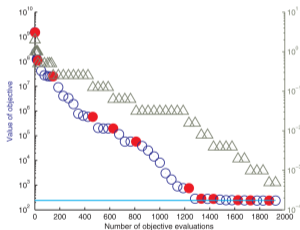
If we only need to use generators of  $T_{\Omega}(\mathbf{x}_k, \alpha_k)$ , why do we need  $\Lambda$ -PSS theory?

# PSS for Linear Constraints

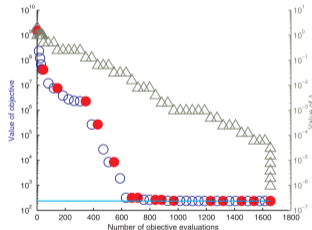
## Why bother?

If we only need to use generators of  $T_{\Omega}(\mathbf{x}_k, \alpha_k)$ , why do we need  $\Lambda$ -PSS theory?

Adding **directions outside the tangent cone** is **practically successful without theoretical justification**—typically (scaled) normal vectors for nearly active constraints.



Tangent generators only



With normal directions

Figures from [Lewis, Shepherd & Torczon, 2007]

## Adding constraint normals

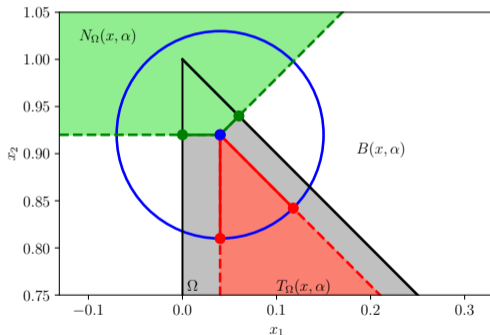
Is (generators of tangent cone) + (nearly active normal vectors) a  $\Lambda$ -PSS?

# PSS for Linear Constraints

## Adding constraint normals

Is (generators of tangent cone) + (nearly active normal vectors) a  $\Lambda$ -PSS?

**No!**



**Example: need (arbitrarily) large coefficients to reach  $v = \text{top corner}$**

## Adding constraint normals

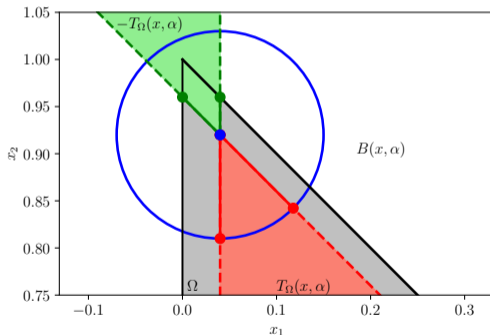
What other vectors could we add to cover the feasible region outside the tangent cone?

# PSS for Linear Constraints

## Adding constraint normals

What other vectors could we add to cover the feasible region outside the tangent cone?

**Use negatives of tangent generators (scaled to be feasible)!**



**Example: use negative of tangent cone generators**

## PSS for Linear Constraints

We can prove that this is a valid  $\Lambda$ -PSS in some settings.

We can prove that this is a valid  $\Lambda$ -PSS in some settings.

### Theorem (LR & Royer, 2026)

*If  $\{\mathbf{a}_j : j \in \mathcal{J}(\mathbf{x}_k, \alpha_k)\}$  is linearly independent, then  $\pm$  generators of tangent cone (scaled to be in  $B(\mathbf{x}_k, \alpha_k) \cap \Omega$ ) is a  $\Lambda$ -PSS with  $\Lambda = |\mathcal{J}| \kappa(A_{\mathcal{J}}) + \sqrt{n - |\mathcal{J}|}$ , where  $A_{\mathcal{J}}$  is the matrix with columns  $\mathbf{a}_j$ .*

We can prove that this is a valid  $\Lambda$ -PSS in some settings.

## Theorem (LR & Royer, 2026)

*If  $\{\mathbf{a}_j : j \in \mathcal{J}(\mathbf{x}_k, \alpha_k)\}$  is linearly independent, then  $\pm$  generators of tangent cone (scaled to be in  $B(\mathbf{x}_k, \alpha_k) \cap \Omega$ ) is a  $\Lambda$ -PSS with  $\Lambda = |\mathcal{J}| \kappa(A_{\mathcal{J}}) + \sqrt{n - |\mathcal{J}|}$ , where  $A_{\mathcal{J}}$  is the matrix with columns  $\mathbf{a}_j$ .*

It is not clear how to guarantee  $\Lambda$ -PSS in the general case, but in practice:

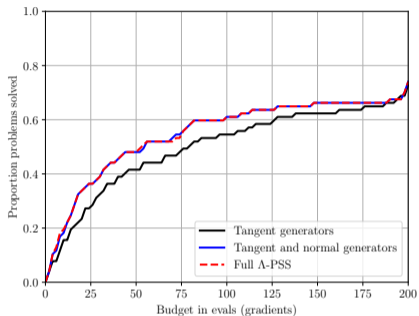
- If  $A_{\mathcal{J}}$  not full rank, use same approach [Abramson et al., 2008]
  - Enumerating generators requires some care e.g. [Fukuda & Prodon, 1996]
- If  $T_{\Omega}(\mathbf{x}_k, \alpha_k) = \{\mathbf{0}\}$ , use (scaled) nearly active normals (guaranteed to be a PSS)
- If  $\pm$  generators of  $T_{\Omega}$  do not (linearly) span  $\mathbb{R}^n$ , augment with poll set in the null space (recursively)

# Numerical Results

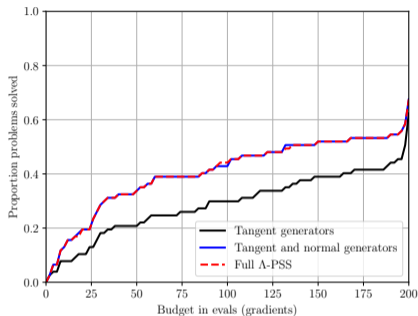
For our testing:

- Compare poll sets from:
  - Existing theory: tangent generators only
  - Existing heuristic approach: tangent generators + nearly active normals
  - Our approach:  $\pm$  tangent generators
- Opportunistic polling (i.e. accept the first poll direction that achieves sufficient decrease)
- Only try extra poll directions if all tangent generators do not achieve sufficient decrease
- Test on 77 bound-constrained and 45 linear inequality constrained CUTEst problems (low dimension,  $n \leq 51$ ) for up to  $200(n + 1)$  objective evaluations

# Numerical Results — Bound Constraints



Low accuracy,  $\tau = 10^{-3}$



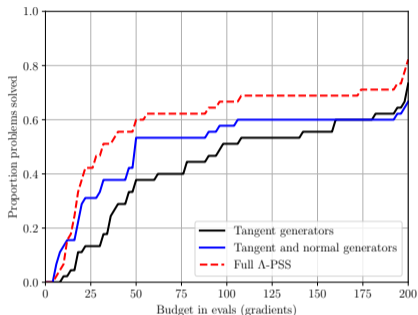
High accuracy,  $\tau = 10^{-6}$

## Bound-constrained problems

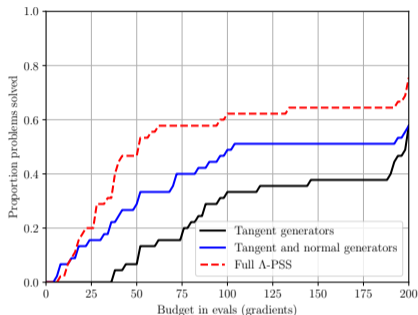
*Data profiles: proportion of problems solved after given number of objective evaluations (units of  $n + 1$ ).*

*Higher is better.*

# Numerical Results — Linear Inequality Constraints



Low accuracy,  $\tau = 10^{-3}$



High accuracy,  $\tau = 10^{-6}$

## Linear inequality constrained problems

Data profiles: proportion of problems solved after given number of objective evaluations (units of  $n + 1$ ).

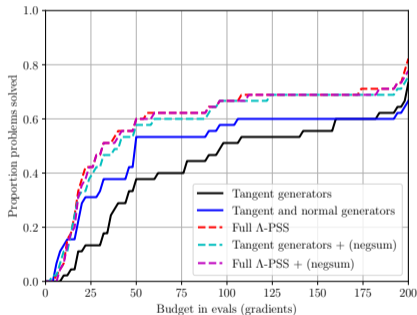
Higher is better.

## Numerical Results — Linear Inequality Constraints

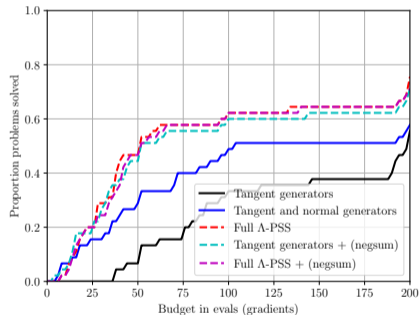
Charles Audet & Stefan Wild at SIOPT: why not use the **sum** of the negative tangent directions as a poll point (to go straight to the corner)?

# Numerical Results — Linear Inequality Constraints

Charles Audet & Stefan Wild at SIOPT: why not use the **sum** of the negative tangent directions as a poll point (to go straight to the corner)?



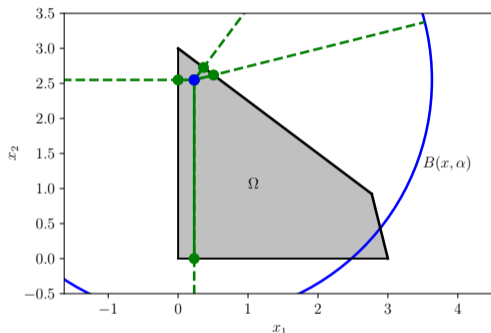
Low accuracy,  $\tau = 10^{-3}$



High accuracy,  $\tau = 10^{-6}$

Linear inequality constrained problems

Practical set construction approach still has room for improvement...



**Example where practical approach can be improved ( $\mathcal{T}_\Omega = \{0\}$ , using active normal directions)**

## Conclusions

- Alternative poll set quality measure,  $\Lambda$ -PSS, generalizes naturally to constrained problems
- Can construct  $\Lambda$ -PSS for bound constraints and some general linear inequality constraints
- Provides some theoretical justification to long-standing practical observation
- Strong numerical performance

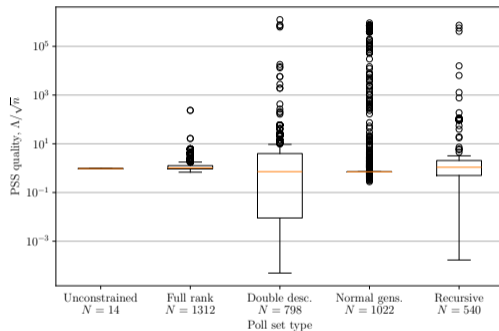
## Future Work

- Full  $\Lambda$ -PSS theory for all possible linear constraints
- Extension to large-scale problems via random embeddings
- Non-polyhedral constraints via curved 'directions' [Jia et al., 2026]

- M. A. ABRAMSON, O. A. BREZHNEVA, J. E. DENNIS, AND R. L. PINGEL, *Pattern search in the presence of degenerate linear constraints*, Optimization Methods and Software, 23 (2008), pp. 297–319.
- C. CARTIS, N. I. M. GOULD, AND P. L. TOINT, *An adaptive cubic regularization algorithm for nonconvex optimization with convex constraints and its function-evaluation complexity*, IMA Journal of Numerical Analysis, 32 (2012), pp. 1662–1695.
- K. FUKUDA AND A. PRODON, *Double description method revisited*, in Combinatorics and Computer Science, G. Goos, J. Hartmanis, J. Leeuwen, M. Deza, R. Euler, and I. Manoussakis, eds., vol. 1120, Springer Berlin Heidelberg, Berlin, Heidelberg, 1996, pp. 91–111.
- S. GRATTON, C. W. ROYER, L. N. VICENTE, AND Z. ZHANG, *Direct search based on probabilistic feasible descent for bound and linearly constrained problems*, Computational Optimization and Applications, 72 (2019), pp. 525–559.
- M. HOUGH AND L. ROBERTS, *Model-based derivative-free methods for convex-constrained optimization*, SIAM Journal on Optimization, 32 (2022), pp. 2552–2579.

- X. JIA, M. LAPUCCI, AND P. MANSUETO, *Projection-based curve pattern search for black-box optimization over smooth convex sets*, Optimization Methods and Software, (2026).
- T. G. KOLDA, R. M. LEWIS, AND V. TORCZON, *Optimization by direct search: New perspectives on some classical and modern methods*, SIAM Review, 45 (2003), pp. 385–482.
- , *Stationarity results for generating set search for linearly constrained optimization*, SIAM Journal on Optimization, 17 (2007), pp. 943–968.
- R. M. LEWIS, A. SHEPHERD, AND V. TORCZON, *Implementing generating set search methods for linearly constrained minimization*, SIAM Journal on Scientific Computing, 29 (2007), pp. 2507–2530.
- L. ROBERTS, *Introduction to interpolation-based optimization*, arXiv preprint arXiv:2510.04473, (2025).
- L. ROBERTS AND C. W. ROYER, *Poll set construction and worst-case complexity for direct search under polyhedral convex constraints*, arXiv preprint arXiv:2605.27814, (2026).
- L. N. TREFETHEN, *Approximation Theory and Approximation Practice*, SIAM, Philadelphia, 2020.

L. N. VICENTE, *Worst case complexity of direct search*, EURO Journal on Computational Optimization, 1 (2013), pp. 143–153.



## Distribution of realized $\Lambda$ for linearly constrained problems