Direct Search for Linearly Constrained Derivative-Free Optimization

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

Lindon Roberts, University of Melbourne (lindon.roberts@unimelb.edu.au)

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

WOMBAT, University of Queensland
26 November 2025

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, *in preparation* (2025).

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

Example Motivation: Survey Construction

Problem: *we want to run a large-scale survey of a population, by picking a subset of people to interview. We want to achieve good survey statistics (e.g. bias, variance).*

Example Motivation: Survey Construction

Problem: *we want to run a large-scale survey of a population, by picking a subset of people to interview. We want to achieve good survey statistics (e.g. bias, variance).*

One continuous formulation of this problem is: survey person i with probability x_i, such that

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \; f(\boldsymbol{x}) \quad \text{(bias, variance, etc.)}, \qquad \text{s.t. } x_i \in [0,1], \; \sum_{i=1}^{n} x_i \leq M,$$

where n is the population and $M \ll n$ is the desired survey size

Example Motivation: Survey Construction

Problem: *we want to run a large-scale survey of a population, by picking a subset of people to interview. We want to achieve good survey statistics (e.g. bias, variance).*

One continuous formulation of this problem is: survey person i with probability x_i, such that

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) \quad \text{(bias, variance, etc.)}, \qquad \text{s.t. } x_i \in [0,1], \sum_{i=1}^{n} x_i \leq M,$$

where n is the population and $M \ll n$ is the desired survey size

Interested in the case where $f(\boldsymbol{x})$ is hard to evaluate, possibly stochastic (e.g. complex Monte Carlo simulation).

Example Motivation: Survey Construction

Problem: *we want to run a large-scale survey of a population, by picking a subset of people to interview. We want to achieve good survey statistics (e.g. bias, variance).*

One continuous formulation of this problem is: survey person i with probability x_i, such that

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) \quad \text{(bias, variance, etc.)}, \qquad \text{s.t. } x_i \in [0,1], \sum_{i=1}^{n} x_i \leq M,$$

where n is the population and $M \ll n$ is the desired survey size

Interested in the case where $f(\boldsymbol{x})$ is hard to evaluate, possibly stochastic (e.g. complex Monte Carlo simulation). This gives problems of the type

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}), \qquad \text{s.t. } A\boldsymbol{x} \leq \boldsymbol{b},$$

where f is black-box, expensive to evaluate ($\rightarrow$ derivatives not readily available).

PhD Position

PhD Position

Fully funded PhD position at UniMelb through OPTIMA to work on this topic with the Australian Bureau of Statistics.

Main topic is derivative-free algorithms for stochastic multi-objective optimization and applications to survey design.

EOI open now, see `OPTIMA_ARC` on LinkedIn or email me for details — student must have the right to study in Australia.

OPTiMA ARC TRAINING CENTRE IN OPTIMISATION TECHNOLOGIES INTEGRATED METHODOLOGIES AND APPLICATIONS

Outline

1. **Unconstrained direct search**
2. Adding convex constraints
3. Specialization to linear constraints
4. Numerical results

Direct Search Methods

Direct search methods are a simple framework for derivative-free optimization (DFO), based purely on sampling function values. [Kolda, Lewis & Torczon, 2003]

Direct Search Methods

Direct search methods are a simple framework for derivative-free optimization (DFO), based purely on sampling function values. [Kolda, Lewis & Torczon, 2003]

Direct Search Iteration

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

Direct Search Methods

Direct search methods are a simple framework for derivative-free optimization (DFO), based purely on sampling function values. [Kolda, Lewis & Torczon, 2003]

Direct Search Iteration

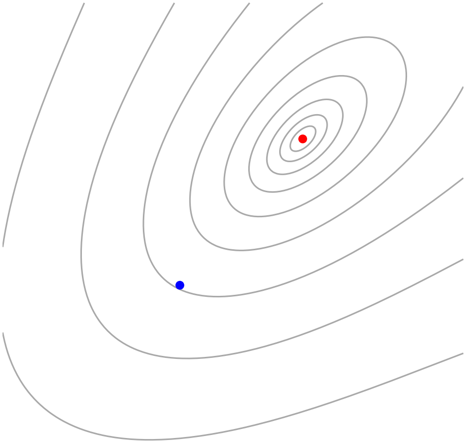
- Given $\boldsymbol{x}_k \in \mathbb{R}^n$ and $\alpha_k > 0$, choose a set $\mathcal{D}_k \subset \mathbb{R}^n$ of p vectors
- If there exists $\boldsymbol{d}_k \in \mathcal{D}_k$ with $f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k) < f(\boldsymbol{x}_k) - \frac{1}{2}\alpha_k^2$
  - Set $\boldsymbol{x}_{k+1}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

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

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

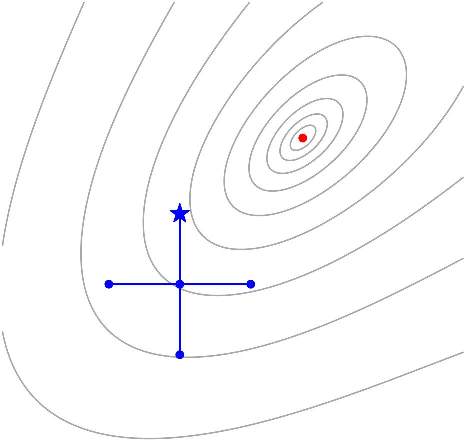
i.e. there is a vector $\boldsymbol{d}$ making a (sufficiently small) acute angle with $-\nabla f(\boldsymbol{x}_k)$.

Example: Direct Search



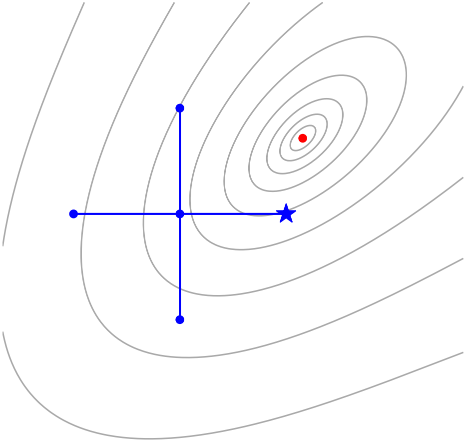
Modified from [Kolda, Lewis & Torczon, 2003]

Example: Direct Search



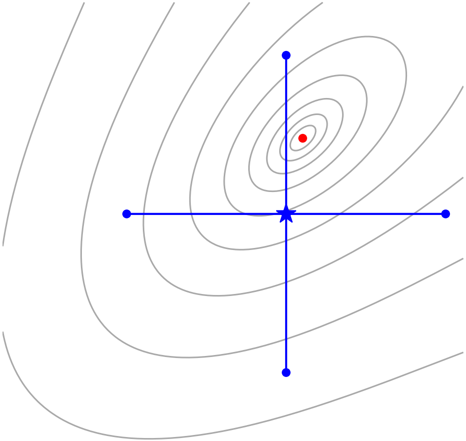
Modified from [Kolda, Lewis & Torczon, 2003]

Example: Direct Search



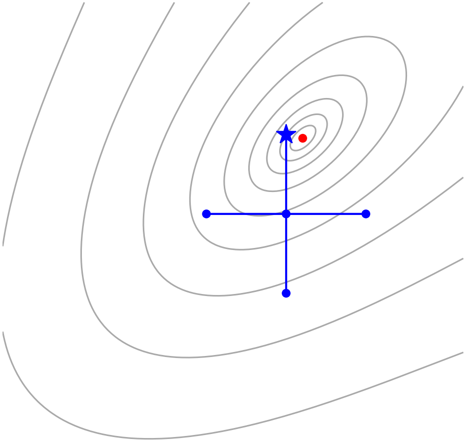
Modified from [Kolda, Lewis & Torczon, 2003]

Example: Direct Search



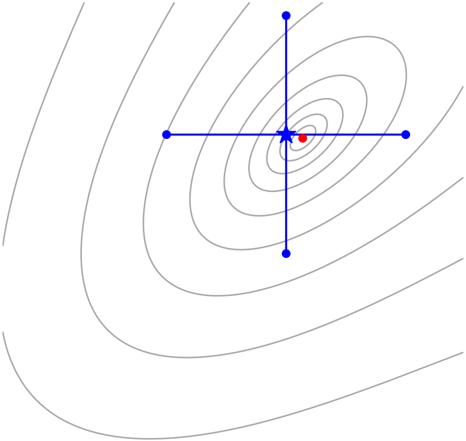
Modified from [Kolda, Lewis & Torczon, 2003]

Example: Direct Search



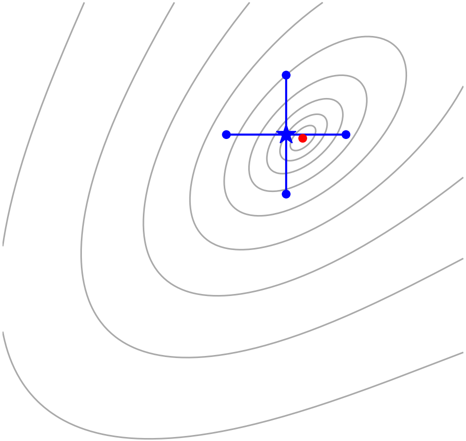
Modified from [Kolda, Lewis & Torczon, 2003]

Example: Direct Search



Modified from [Kolda, Lewis & Torczon, 2003]

Example: Direct Search



Modified from [Kolda, Lewis & Torczon, 2003]

Direct Search: Convergence Theory

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

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

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

Direct Search: Convergence Theory

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

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

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

Theorem (Vicente, 2013)

If f has Lipschitz continuous gradients, $\|\boldsymbol{d}\|_2 \leq d_{\max}$ for all $\boldsymbol{d}\in\mathcal{D}_k$ and $\mathrm{cm}(\mathcal{D}_k) \geq \kappa > 0$ for all k, then $\|\nabla f(\boldsymbol{x}_k)\|_2 \leq \epsilon$ 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 $\mathcal{D}_k$:

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

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

Theorem (Vicente, 2013)

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

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

Direct Search: Convergence Theory

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

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

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

Theorem (Vicente, 2013)

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

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

Note: $\mathcal{O}(n^2)$ dependency cannot be improved without randomization.

[Dodangeh et al., 2016], [Gratton et al., 2015], [LR & Royer, 2023]

An alternative perspective

Link 1: A classical result is:

Theorem (Davis, 1954)

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

An alternative perspective

Link 1: A classical result is:

Theorem (Davis, 1954)

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

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

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

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

This corresponds to the Lebesgue measure in approximation theory.

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

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

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

Theorem (LR & Royer, 2025)

(a) *If $\mathcal{D}$ is a Λ-PSS and and $\|\boldsymbol{d}_i\|_2 \leq d_{\max}\alpha$, then $\mathrm{cm}(\mathcal{D}) \geq 1/(d_{\max}\Lambda)$*

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

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

Outline

1. Unconstrained direct search
2. **Adding convex constraints**
3. Specialization to linear constraints
4. Numerical results

Convex Constraints

Now, consider the constrained problem

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

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

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

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

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

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

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

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

Issue: no natural generalization of $\mathrm{cm}(\mathcal{D})$ to the constrained setting.

Constrained Positive Spanning Sets

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

Definition

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

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

Constrained Positive Spanning Sets

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

Definition

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

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, 2025)

If f has Lipschitz continuous gradients, $\|\boldsymbol{d}\|_2 \leq d_{\max}\alpha_k$ for all $\boldsymbol{d} \in \mathcal{D}_k$ and $\mathcal{D}_k$ is a Λ-PSS for $B(\boldsymbol{x}_k, \alpha_k) \cap \Omega$ for all k, then $\pi(\boldsymbol{x}_k) \leq \epsilon$ 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, 2025)

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

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

Constrained Positive Spanning Sets

This enables a very simple worst-case complexity result.

Theorem (LR & Royer, 2025)

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

Key new idea: by definition, $\pi(\boldsymbol{x}_k) = -\nabla f(\boldsymbol{x}_k)^T \boldsymbol{v}_k$ for some $\|\boldsymbol{v}_k\|_2 \leq 1$ with $\boldsymbol{x}_k + \boldsymbol{v}_k \in \Omega$, invoke Λ-PSS property (or $\alpha_k \boldsymbol{v}_k$ 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 Λ-PSS sets for constrained problems?

Bound Constraints

Problem

How do we construct Λ-PSS sets for constrained problems?

This is easy for bound-constrained problems.

Bound Constraints

Problem

How do we construct Λ-PSS sets for constrained problems?

This is easy for bound-constrained problems.

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

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

Bound Constraints

Problem

How do we construct Λ-PSS sets for constrained problems?

This is easy for bound-constrained problems.

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

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

Theorem (LR & Royer, 2025)

$\mathcal{D}$ is a Λ-PSS with $\Lambda = n$.

Bound Constraints

Problem

How do we construct Λ-PSS sets for constrained problems?

This is easy for bound-constrained problems.

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

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

Theorem (LR & Royer, 2025)

$\mathcal{D}$ is a Λ-PSS with $\Lambda = n$.

This leads to $\mathcal{O}(n^2\epsilon^{-2})$ iterations or $\mathcal{O}(n^3\epsilon^{-2})$ evaluations to achieve first-order optimality ϵ (worse than unconstrained case, but usually n not too large).

Outline

1. Unconstrained direct search
2. Adding convex constraints
3. **Specialization to linear constraints**
4. Numerical results

Linearly Constrained Problems

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

How is linearly constrained direct search currently performed?

Linearly Constrained Problems

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

How is linearly constrained direct search currently performed?

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

$$j \in \mathcal{J}(\boldsymbol{x}_k, \alpha_k) \qquad \Longleftrightarrow \qquad b_j - \alpha_k \|\boldsymbol{a}_j\|_2^2 \le \boldsymbol{a}_j^T \boldsymbol{x}_k \le b_j$$

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

Linearly Constrained Problems

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

How is linearly constrained direct search currently performed?

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

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

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

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

Linearly Constrained Problems

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

How is linearly constrained direct search currently performed?

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

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

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

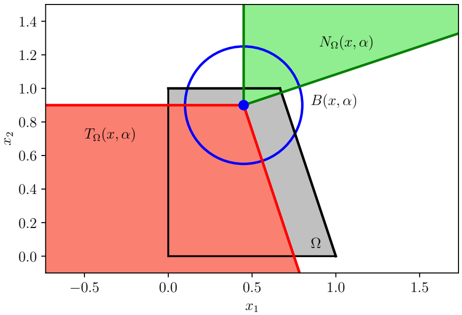
The approximate normal cone $N_\Omega(\boldsymbol{x}_k, \alpha_k)$ is generated by $\boldsymbol{a}_j$ for $j \in \mathcal{J}(\boldsymbol{x}_k, \alpha_k)$.

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

[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 $\boldsymbol{x} = [0.45, 0.9]$:

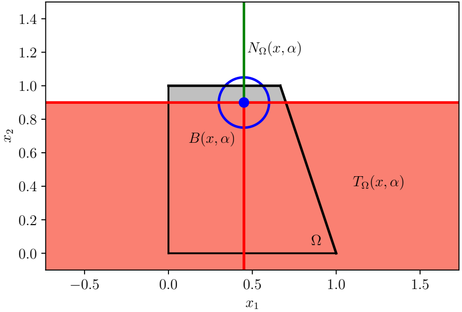


(a) $\alpha = 0.35$, nearly active constraints are $x_2 \leq 1$ and $3x_1 + x_2 \leq 3$

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 $\boldsymbol{x} = [0.45, 0.9]$:



(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

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

such that

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

Linearly Constrained Problems

In existing theory, we choose the poll set such that

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

such that

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

Theorem (Gratton et al., 2019)

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

Linearly Constrained Problems

In existing theory, we choose the poll set such that

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

such that

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

Theorem (Gratton et al., 2019)

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

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

PSS for Linear Constraints

Why bother?

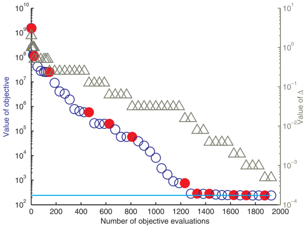
If we only need to use generators of $T_\Omega(\boldsymbol{x}_k, \alpha_k)$, why do we need Λ-PSS theory?

PSS for Linear Constraints

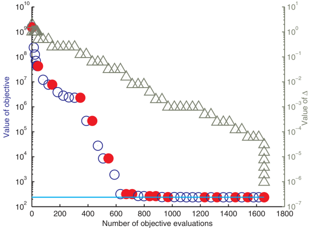
Why bother?

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

Adding directions outside the tangent cone is practically successful (but no theory)—typically (scaled) normal vectors for nearly active constraints.



Tangent generators only



With normal directions

Figures from [Lewis, Shepherd & Torczon, 2007]

PSS for Linear Constraints

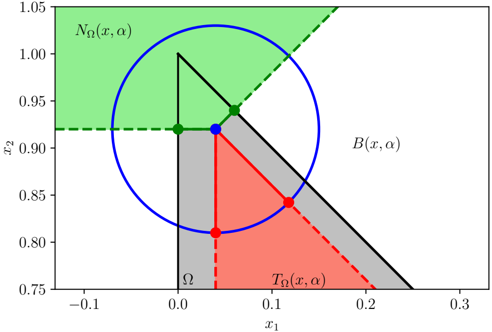
Adding constraint normals

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

Adding constraint normals

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

No!



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

PSS for Linear Constraints

Adding constraint normals

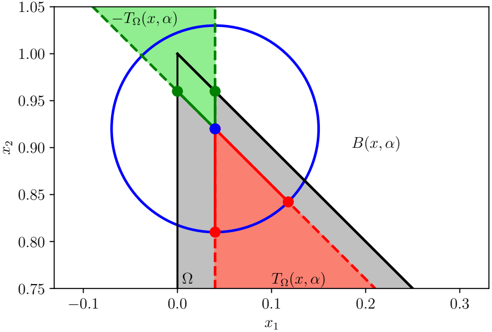
What other vectors could we add to cover the feasible region outside the tangent cone?

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 Λ-PSS in some settings.

PSS for Linear Constraints

We can prove that this is a valid Λ-PSS in some settings.

Theorem (LR & Royer, 2025)

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

PSS for Linear Constraints

We can prove that this is a valid Λ-PSS in some settings.

Theorem (LR & Royer, 2025)

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

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

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

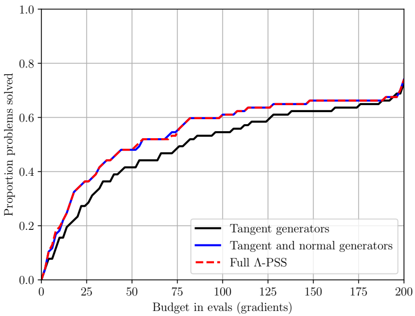
Outline

Numerical Results

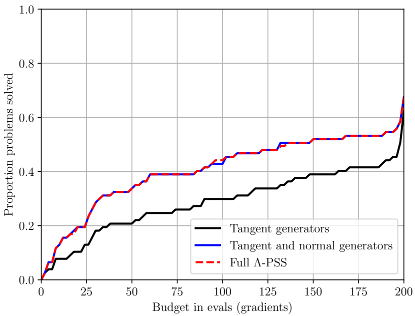
For our testing:

- Compare poll sets from existing theory (tangent generators), existing heuristic approach (tangent generators + nearly active normals) with our approach ($\pm$ tangent generators)
- 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, $\tau = 10^{-3}$



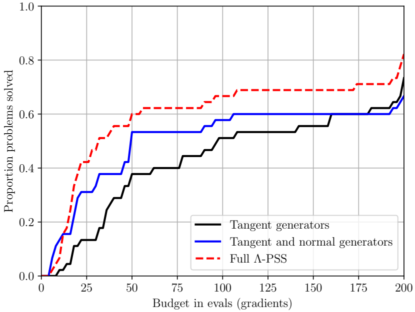
High accuracy, $\tau = 10^{-6}$

Bound-constrained problems

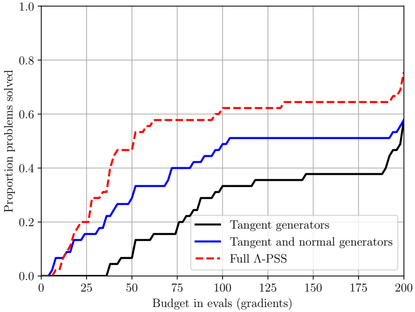
Data profiles: proportion of problems solved after given number of objective evaluations (units of $n+1$).

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.

Conclusions & Future Work

Conclusions

- Alternative poll set quality measure, Λ-PSS, generalizes naturally to constrained problems
- Can construct Λ-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 Λ-PSS theory for all possible linear constraints
- Extension to large-scale problems via random embeddings

References i

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.

M. Dodangeh, L. N. Vicente, and Z. Zhang, *On the optimal order of worst case complexity of direct search*, Optimization Letters, 10 (2016), pp. 699–708.

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 descent*, SIAM Journal on Optimization, 25 (2015), pp. 1515–1541.

———, *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.

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, *Direct search based on probabilistic descent in reduced spaces*, SIAM Journal on Optimization, 33 (2023), pp. 3057–3082.

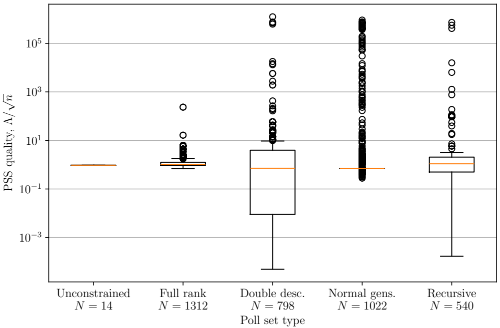
———, *Poll set construction and worst-case complexity for direct search under polyhedral convex constraints*, in preparation, (2025).

L. N. Trefethen, *Approximation Theory and Approximation Practice*, SIAM, Philadelphia, 2020.

References iii

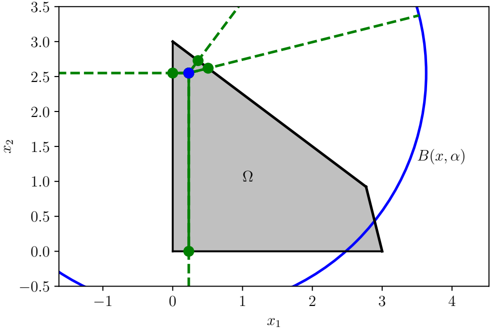
L. N. Vicente, *Worst case complexity of direct search*, EURO Journal on Computational Optimization, 1 (2013), pp. 143–153.

Practical Construction Quality



Distribution of realized Λ for linearly constrained problems

Practical Construction Quality



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