Subspace Modeling and Random Subspace Trust-region Methods in Derivative-free Optimization

Yiwen Chen

Department of Mathematics
University of British Columbia

NOVA Math Seminar, Portugal

July 23, 2026

Joint works with Warren Hare and Amy Wiebe

Outline

The problem: optimize a blackbox

Goal

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

where evaluating $f(x)$ is possible, but derivative information is unavailable or unreliable

Typical situation

Evaluation of f may be a simulation, experiment, or legacy code call

Example: optimizing an expensive experiment

Inputs

$$x = \begin{bmatrix} \text{temperature} \\ \text{pressure} \\ \text{reaction time} \end{bmatrix}$$

Output

$f(x) =$ measured impurity level

$\min_x \ f(x)$ subject to safety and equipment limits

Derivative-free optimization (DFO)

Derivative-free optimization

Derivative-free optimization is the mathematical study of optimization algorithms that do not use derivatives

$$x \longmapsto f(x), \qquad \nabla f(x) \text{ unavailable to the algorithm}$$

Important distinction

It does not mean that derivatives do not exist; it means the algorithm does not access them

Two categories of DFO methods

Direct search methods

- Maintain an incumbent solution and check a finite number of trial points for potential decrease
- E.g., Coordinate Search, Mesh Adaptive Direct Search

Model-based methods

- Use function values to build an approximation model of the objective
- Use the model to guide future iterations

This talk focuses on model-based methods

How model-based DFO constructs models

Definition

For a given function $f(x)$ and set $Y = \{y^0, \ldots, y^s\}$, a polynomial $m(x)$ is a polynomial interpolation model of $f(x)$ if

$$m(y^i) = f(y^i), \quad i = 0, \ldots, s$$

Note: In practice, $m(x)$ is determined by finding $\alpha_0, \ldots, \alpha_t$ such that

$$m(y^i) = \sum_{j=0}^{t} \alpha_j \phi_j(y^i) = f(y^i), \quad i = 0, \ldots, s$$

where the set of functions $\phi = \{\phi_0, \ldots, \phi_t\}$ is a polynomial basis

Linear interpolation model: the first-order model

Definition

Let $\phi = \{1, x_1, x_2, \ldots, x_n\}$ and $Y = \{y^0, \ldots, y^s\} \subseteq \mathbb{R}^n$
A linear interpolation model $m(x)$ is determined by finding $\alpha_0, \ldots, \alpha_n$ such that

$$m(y^i) = \alpha_0 + \alpha_1 y_1^i + \cdots + \alpha_n y_n^i = \begin{bmatrix} 1 & (y^i)^\top \end{bmatrix} \alpha = f(y^i), \quad i = 0, \ldots, s$$

where y_j^i is the j-th element of y^i

Note: If $s + 1 = n + 1$ and the system has full rank, then $m(x)$ is called a determined linear interpolation model

Quadratic interpolation: more info, but more expensive

Definition

Let $\phi = \{1, x_1, x_2, \ldots, x_n, \frac{x_1^2}{2}, x_1x_2, \ldots, x_{n-1}x_n, \frac{x_n^2}{2}\}$ and $Y = \{y^0, \ldots, y^s\} \subseteq \mathbb{R}^n$

A quadratic interpolation model $m(x)$ is determined by finding $\alpha_0, \ldots, \alpha_{\frac{(n+1)(n+2)}{2}-1}$ such that

$$m(y^i) = \alpha_0 + \alpha_1 y_1^i + \cdots + \alpha_n y_n^i + \alpha_{n+1}\frac{(y_1^i)^2}{2} + \cdots + \alpha_{\frac{(n+1)(n+2)}{2}-1}\frac{(y_n^i)^2}{2} = f(y^i), \quad i = 0, \ldots, s$$

Note: If $s + 1 = \frac{(n+1)(n+2)}{2}$ and the system has full rank, then $m(x)$ is called a determined quadratic interpolation model

The bottleneck: scalability

The scaling bottleneck

A determined quadratic interpolation model in $\mathbb{R}^n$ needs

$$\frac{(n+1)(n+2)}{2} \quad \text{sample points}$$

n	1	10	100	1000
sample points	3	66	5151	501501

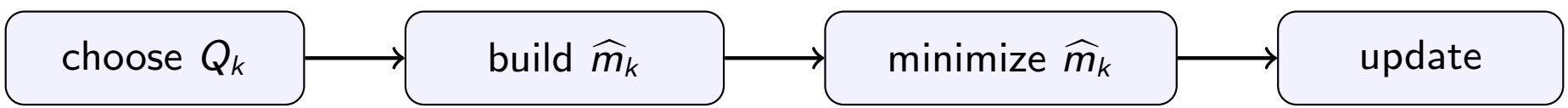
Full-space quadratic models become impractical for high-dimensional problems ($n \approx 1000$)

The main idea: model only in subspaces

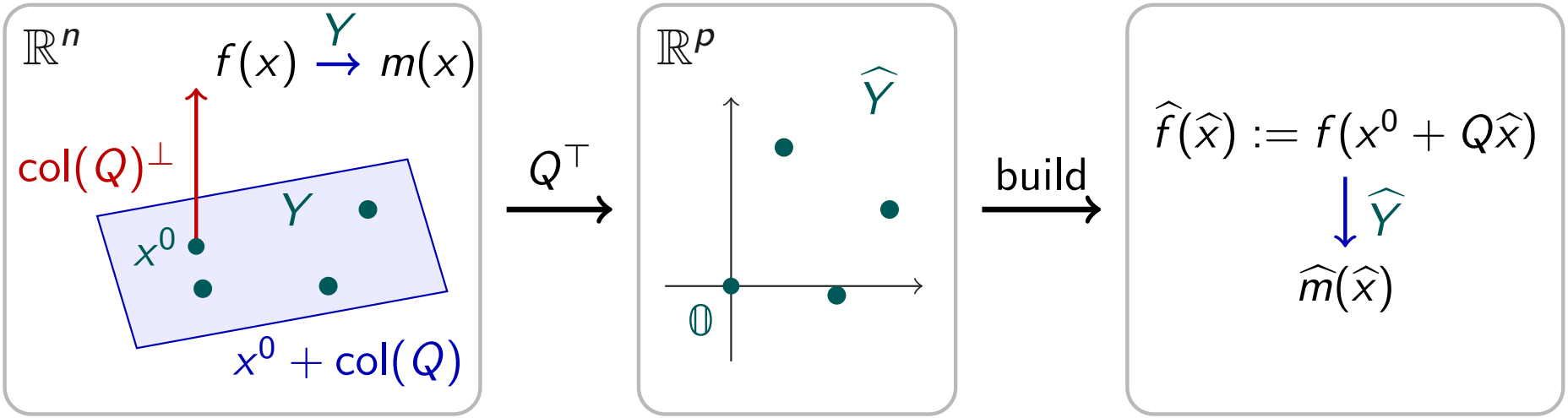
Idea

At iteration k, replace a hard n-dimensional modeling problem with a lower-dimensional one on

$$x_k + Q_k \mathbb{R}^p, \qquad p \ll n$$



Two equivalent views of subspace models



Lift and equivalence (Chen 2026)

For many widely-used linear/quadratic models, we have

$$\nabla m(x^0) = Q\nabla \widehat{m}(\mathbb{0}), \quad \nabla^2 m(x^0) = Q\nabla^2 \widehat{m}(\mathbb{0})Q^\top,$$

and

$$m(x) = \widehat{m}(\widehat{x}), \quad \text{for all } x \in \left(x^0 + Q\widehat{x}\right) + \mathrm{col}(Q)^\perp$$

Outline

Model-based trust-region (MBTR) algorithm

for $k = 0, 1, \ldots$ **do**

Construct a model m_k in $\mathbb{R}^n$:

$$m_k(s) = f(x_k) + g_k^\top s + \frac{1}{2} s^\top H_k s$$

Approximately solve the trust-region subproblem in $\mathbb{R}^n$:

$$s_k \approx \operatorname*{argmin}_{s \in \mathbb{R}^n} m_k(s), \quad s.t. \ \|s\| \leq \Delta_k$$

Evaluate $f(x_k + s_k)$ and apply descent ratio test

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{m_k(\mathbb{0}) - m_k(s_k)} = \frac{\textit{true decrease}}{\textit{predicted decrease}}$$

Accept/reject step based on ρ_k and update trust region radius

Random subspace MBTR algorithm [Cartis, Roberts, 2023]

for $k = 0, 1, \ldots$ **do**

Define an affine subspace $x_k + \mathrm{col}(Q_k)$ by selecting $Q_k \in \mathbb{R}^{n \times p}$

Construct a model $\widehat{m}_k$ in $\mathbb{R}^p$

Approximately solve the trust-region subproblem in $\mathbb{R}^p$:

$$\widehat{s}_k \approx \underset{\widehat{s} \in \mathbb{R}^p}{\operatorname{argmin}} \, \widehat{m}_k(\widehat{s}), \quad s.t. \ \ \|\widehat{s}\| \leq \Delta_k$$

and calculate the corresponding step $s_k \in \mathbb{R}^n$

Evaluate $f(x_k + s_k)$ and apply descent ratio test

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{\widehat{m}_k(\mathbb{0}) - \widehat{m}_k(\widehat{s}_k)} = \frac{\textit{true decrease}}{\textit{predicted decrease}}$$

Accept/reject step based on ρ_k and update trust region radius

Models with "good enough" accuracy

Definition

A model $m : \mathbb{R}^n \to \mathbb{R}$ is fully linear in $B(x, \Delta)$ if there exist $\kappa_f, \kappa_g > 0$ s.t. for all $s \in \mathbb{R}^n$ with $\|s\| \leq \Delta$,

$$|f(x+s) - m(x+s)| \leq \kappa_f \Delta^2$$
$$\|\nabla f(x+s) - \nabla m(x+s)\| \leq \kappa_g \Delta$$

Taylor's theorem

Let $f \in \mathcal{C}^{1+}$ in $B(x, \Delta)$ with constant $L_{\nabla f}$
Then, for all $s \in \mathbb{R}^n$ with $\|s\| \leq \Delta$

$$\left| f(x+s) - f(x) - \nabla f(x)^\top s \right| \leq \frac{1}{2} L_{\nabla f} \Delta^2$$
$$\|\nabla f(x+s) - \nabla f(x)\| \leq L_{\nabla f} \Delta$$

Note: Determined linear interpolation model is fully linear!

Model accuracy is only needed in the searched subspace

Definition

A model $m : \mathbb{R}^n \to \mathbb{R}$ is fully linear in $B(x, \Delta)$ if there exist $\kappa_f, \kappa_g > 0$ s.t. for all $s \in \mathbb{R}^n$ with $\|s\| \leq \Delta$,

$$|f(x+s) - m(x+s)| \leq \kappa_f \Delta^2$$
$$\|\nabla f(x+s) - \nabla m(x+s)\| \leq \kappa_g \Delta$$

Definition (Cartis, Roberts, 2023)

Let $Q \in \mathbb{R}^{n \times p}$. A model $\widehat{m} : \mathbb{R}^p \to \mathbb{R}$ is Q-fully linear in $B(x, \Delta)$ if there exist $\kappa_f, \kappa_g > 0$ s.t. for all $\widehat{s} \in \mathbb{R}^p$ with $\|\widehat{s}\| \leq \Delta$,

$$|f(x + Q\widehat{s}) - \widehat{m}(\widehat{s})| \leq \kappa_f \Delta^2$$
$$\left\| Q^\top \nabla f(x + Q\widehat{s}) - \nabla \widehat{m}(\widehat{s}) \right\| \leq \kappa_g \Delta$$

Quadratic models? $\Rightarrow$ QARSTA: Quadratic Approximation Random Subspace Trust-region Algorithm [Chen, Hare, Wiebe, 2024]

Quadratic accuracy: captures subspace curvature info

Definition

A model $m : \mathbb{R}^n \to \mathbb{R}$ is fully quadratic in $B(x, \Delta)$ if there exist $\kappa_f, \kappa_g, \kappa_h > 0$ s.t. for all $s \in \mathbb{R}^n$ with $\|s\| \leq \Delta$,

$$|f(x+s) - m(x+s)| \leq \kappa_f \Delta^3$$
$$\|\nabla f(x+s) - \nabla m(x+s)\| \leq \kappa_g \Delta^2$$
$$\left\|\nabla^2 f(x+s) - \nabla^2 m(x+s)\right\| \leq \kappa_h \Delta$$

Definition (Chen, Hare, Wiebe, 2024)

Let $Q \in \mathbb{R}^{n \times p}$. A model $\widehat{m} : \mathbb{R}^p \to \mathbb{R}$ is Q-fully quadratic in $B(x, \Delta)$ if there exist $\kappa_f, \kappa_g, \kappa_h > 0$ s.t. for all $\widehat{s} \in \mathbb{R}^p$ with $\|\widehat{s}\| \leq \Delta$,

$$|f(x + Q\widehat{s}) - \widehat{m}(\widehat{s})| \leq \kappa_f \Delta^3$$
$$\left\|Q^\top \nabla f(x + Q\widehat{s}) - \nabla \widehat{m}(\widehat{s})\right\| \leq \kappa_g \Delta^2$$
$$\left\|Q^\top \nabla^2 f(x + Q\widehat{s}) Q - \nabla^2 \widehat{m}(\widehat{s})\right\| \leq \kappa_h \Delta$$

Constructing Q-fully linear/quadratic models

Let D consist of sample directions and write $D = QR$, where $Q^\top Q = I_p$
Denote $\overline{\operatorname{diam}}(R) := \max_{r_i \in R} \|r_i\|$

Theorem

If $f \in \mathcal{C}^{1+}$, D has full column rank, and

$$\left\|\overline{R}^{-1}\right\| \quad \textit{is uniformly bounded, where } \overline{R} := R/\overline{\operatorname{diam}}(R),$$

then the subspace linear interpolation model is Q-fully linear in $B(x^0; \overline{\operatorname{diam}}(R))$

Constructing Q-fully linear/quadratic models

Let D consist of sample directions and write $D = QR$, where $Q^\top Q = I_p$
Denote $\overline{\operatorname{diam}}(R) := \max_{r_i \in R} \|r_i\|$

Theorem

If $f \in \mathcal{C}^{1+}$, D has full column rank, and

$$\left\|\overline{R}^{-1}\right\| \quad \textit{is uniformly bounded, where } \overline{R} := R/\overline{\operatorname{diam}}(R),$$

then the subspace linear interpolation model is Q-fully linear in $B(x^0; \overline{\operatorname{diam}}(R))$

Theorem (Chen, Hare, Wiebe, 2024)

If we replace $f \in \mathcal{C}^{1+}$ by $f \in \mathcal{C}^{2+}$ in the theorem above, then the subspace quadratic interpolation model is Q-fully quadratic in $B(x^0; \overline{\operatorname{diam}}(R))$

Random subspaces: when do they preserve descent info?

Let $x + \mathrm{col}(D)$ be the affine subspace

Definition (Cartis, Roberts, 2023)

Let $\alpha \in (0, 1)$. We say that $D \in \mathbb{R}^{n \times p}$ is α-well-aligned for f at x if

$$\left\| D^\top \nabla f(x) \right\| \geq \alpha \left\| \nabla f(x) \right\|$$

Random subspaces: when do they preserve descent info?

Let $x + \mathrm{col}(D)$ be the affine subspace

Definition (Cartis, Roberts, 2023)

Let $\alpha \in (0, 1)$. We say that $D \in \mathbb{R}^{n \times p}$ is α-well-aligned for f at x if

$$\left\| D^\top \nabla f(x) \right\| \geq \alpha \left\| \nabla f(x) \right\|$$

Theorem (Dzahini, Wild 2024)

(Idea: Johnson–Lindenstrauss Lemma)

Suppose $\alpha, \delta \in (0, 1)$, $p \geq 4(1 - \alpha)^{-2} \ln(1/\delta)$, *and i.i.d.* $D_{ij} \sim \mathcal{N}(0, 1/p)$. *Then,*

$$\mathbb{P}\left[D \text{ is } \alpha\text{-well-aligned for } f \text{ at } x\right] \geq 1 - \delta$$

Beyond pure randomness: reusing past information

Let D consist of sample directions and write

$$D = [D^U \; D^R] \in \mathbb{R}^{n \times p}$$

Sample set management idea

- Reuse some directions D^U from previous sample points
- Add at least p_{rand} new random directions D^R
- Choose D^U and D^R so that $\|\overline{R}^{-1}\|$ stays uniformly bounded

Theorem (Chen, Hare, Wiebe, 2024)

When adding a new direction of length Δ, the best choice for the geometry measure $\|\overline{R}^{-1}\|$ is a direction orthogonal to the existing columns.

Reused directions still preserve descent information

Theorem (Chen, Hare, Wiebe, 2024)

Let $\alpha, \delta \in (0,1)$ *and suppose* $p_{\text{rand}} \geq 4(1-\alpha)^{-2}\ln(1/\delta)$
If the new random directions are generated orthogonally to the reused directions, then there exists $\alpha_D > 0$ *such that*

$$\mathbb{P}\left[D \text{ is } \alpha_D\text{-well-aligned for } f \text{ at } x\right] \geq 1-\delta$$

Even with reused sample directions, the subspace is still well-aligned with high probability

Numerical results

What the plots compare

- Linear vs. quadratic subspace models
- Different values of p and p_{rand}

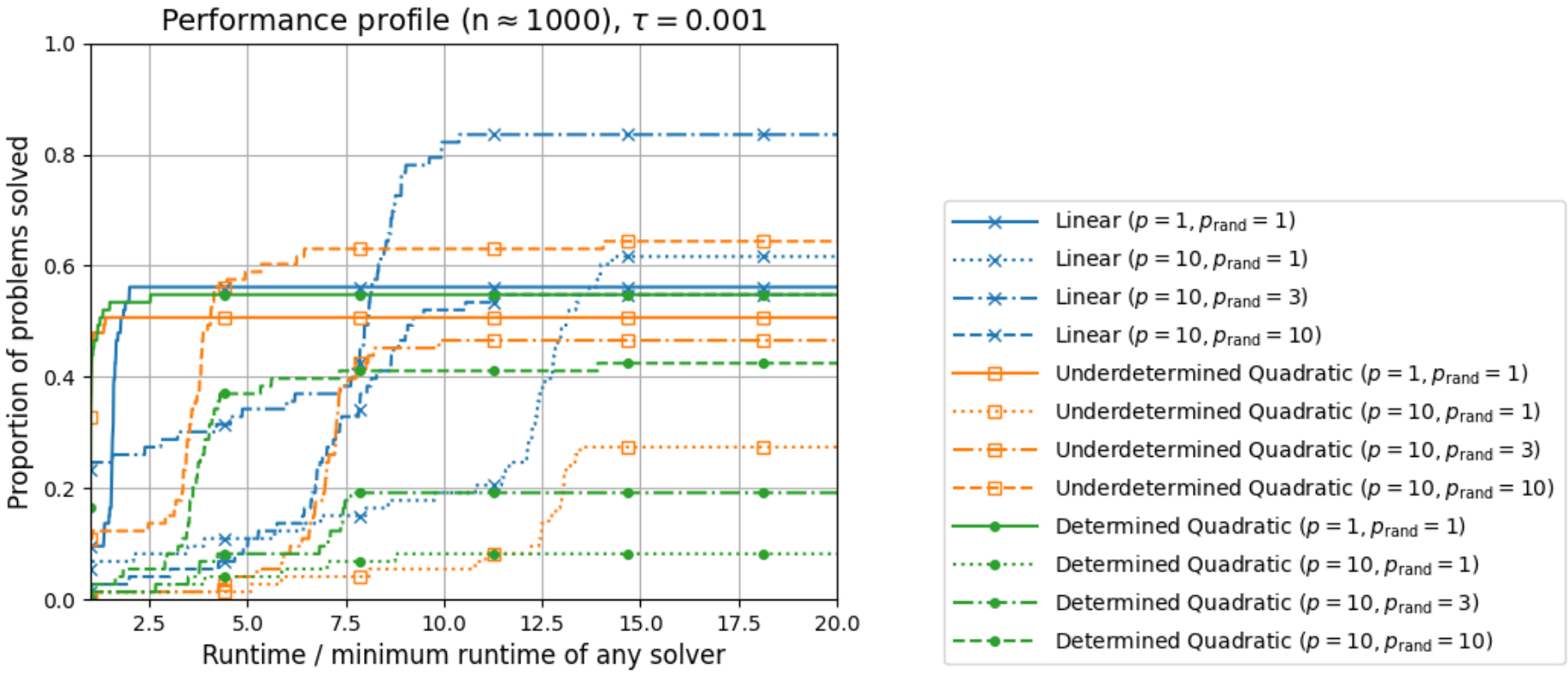
Runtime

Does the additional quadratic modelling cost pay off?

Function evaluations

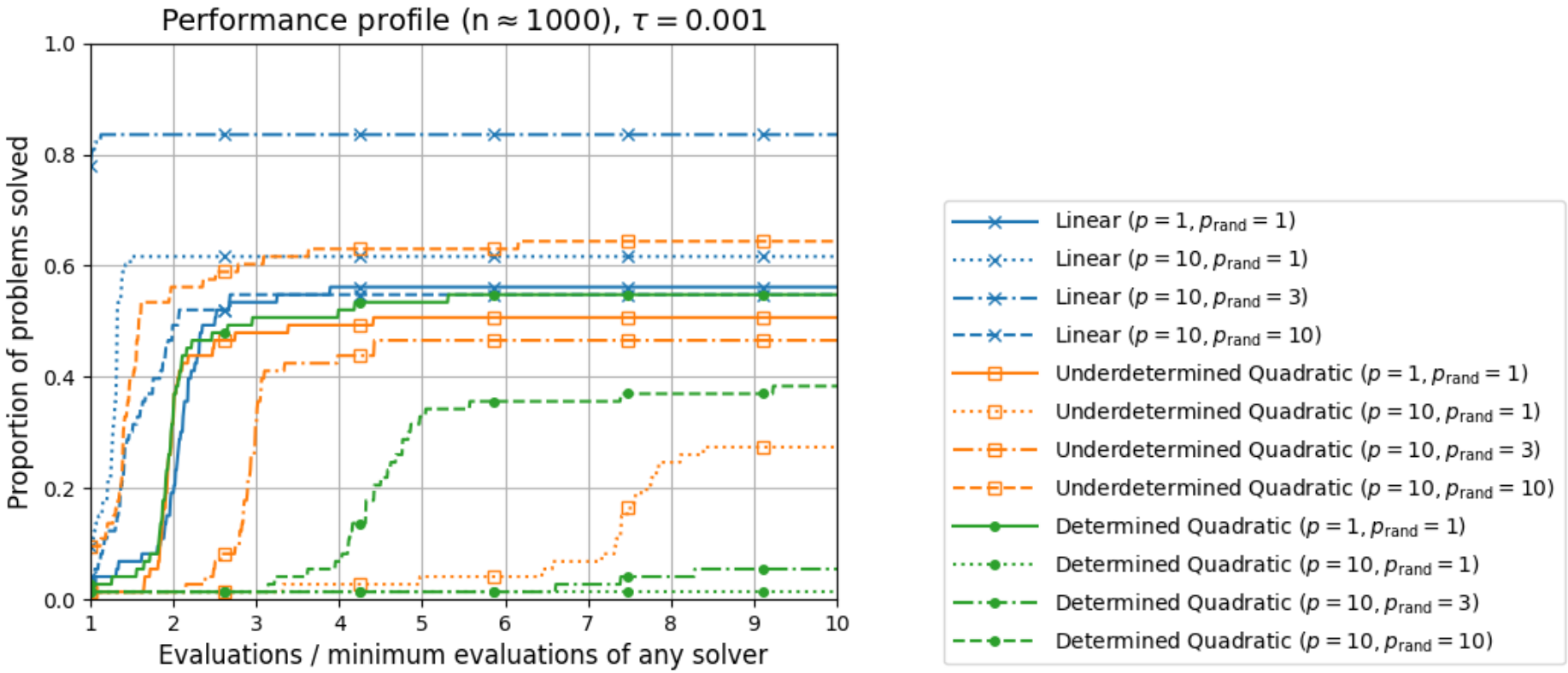
Does higher model accuracy reduce expensive blackbox calls?

Comparing linear and quadratic models based on runtime



Success is defined as finding x_k such that
$f(x_k) \leq f(x^*) + \tau(f(x_0) - f(x^*))$ in less than $100(n+1)$ f-evals

Comparing linear and quadratic models based on f-evals



Success is defined as finding x_k such that
$f(x_k) \leq f(x^*) + \tau(f(x_0) - f(x^*))$ in less than $100(n+1)$ f-evals

Outline

What changes when constraints are present?

Main story

CLARSTA is a random-subspace trust-region framework for

$$\min_{x \in C} f(x)$$

where C is convex, closed, and has a nonempty interior

From QARSTA to CLARSTA

for $k = 0, 1, \ldots$ **do**

Define an affine subspace $x_k + \mathrm{col}(Q_k)$ by selecting $Q_k \in \mathbb{R}^{n \times p}$

Construct a model $\widehat{m}_k$ in $\mathbb{R}^p$

Approximately solve the trust-region subproblem in $\mathbb{R}^p$:

$$\widehat{s}_k \approx \underset{\widehat{s} \in \mathbb{R}^p}{\operatorname{argmin}} \, \widehat{m}_k(\widehat{s}), \quad s.t. \ \ \|\widehat{s}\| \leq \Delta_k$$

and calculate the corresponding step $s_k \in \mathbb{R}^n$

Evaluate $f(x_k + s_k)$ and calculate descent ratio

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{\widehat{m}_k(\mathbb{0}) - \widehat{m}_k(\widehat{s}_k)} = \frac{\textit{true decrease}}{\textit{predicted decrease}}$$

Accept/reject step based on ρ_k and update trust region radius

CLARSTA: random subspace trust regions with feasibility

for $k = 0, 1, \ldots$ **do**

Define an affine subspace $x_k + \mathrm{col}(Q_k)$ by *selecting* $Q_k \in \mathbb{R}^{n \times p}$

Construct a model $\widehat{m}_k$ in $\mathbb{R}^p$

Approximately solve the trust-region subproblem in $\mathbb{R}^p$:

$$\widehat{s}_k \approx \underset{\widehat{s} \in Q_k^\top (C - x_k)}{\operatorname{argmin}} \ \widehat{m}_k(\widehat{s}), \quad s.t. \ \|\widehat{s}\| \leq \Delta_k$$

and calculate the corresponding feasible step $s_k \in \mathbb{R}^n$

Evaluate $f(x_k + s_k)$ and calculate descent ratio

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{\widehat{m}_k(\mathbb{0}) - \widehat{m}_k(\widehat{s}_k)} = \frac{\textit{true decrease}}{\textit{predicted decrease}}$$

Accept/reject step based on ρ_k and update trust region radius

Accuracy is only needed on projected feasible directions

Model class	Accuracy region	How to build it
Classic fully linear	$B(x^0; \Delta)$	Full-space linear interpolation
Q-fully linear	$B(x^0; \Delta) \cap \left(x^0 + \mathrm{col}(Q)\right)$	Subspace linear interpolation
(C, Q)-fully linear	$B(x^0; \Delta) \cap \mathrm{proj}_{x^0+\mathrm{col}(Q)}(C)$	Subspace linear interpolation with *feasible geometry*

Fully/Q-fully linear bounds can blow up with constraints

Problem (Larson, Menickelly, Wild 2019)

"For a fixed value of κ, it may be impossible to obtain a κ-fully linear model using only feasible points"

Constrained example (Hough, Roberts 2022)

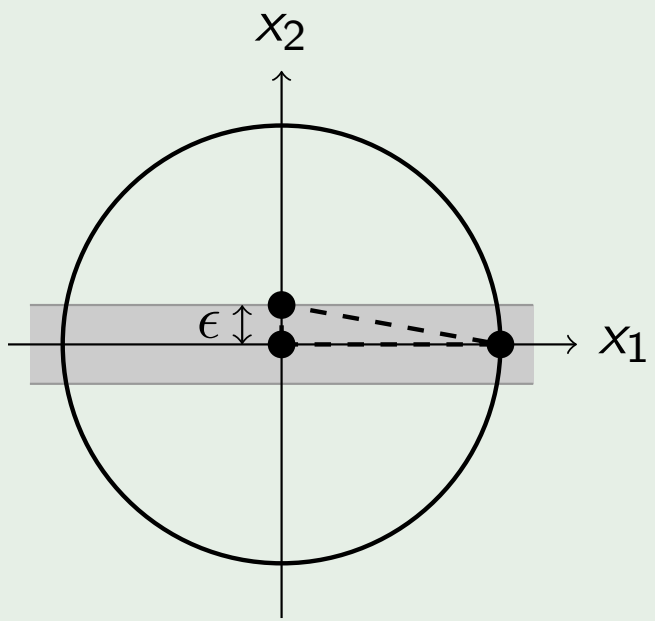
Consider

$$C = \{(x_1, x_2) : |x_2| \leq \epsilon\} \subseteq \mathbb{R}^2$$

and sample set

$$Y = \{\mathbb{0},\ (1,0),\ (0,\epsilon)\} \subseteq B(\mathbb{0}, 1) \cap C$$

Then, $\|\overline{R}^{-1}\| \sim \epsilon^{-1}$



A geometry measure that respects the feasible set

Let D consist of sample directions and write $D = QR$, where $Q^\top Q = I_p$

Feasible geometry measure: matrix spectral norm w.r.t. $Q^\top(C - x^0)$

For any matrix $M \in \mathbb{R}^{m \times p}$, define

$$\|M\|_{x^0,C,D} := \frac{1}{\min\{\overline{\operatorname{diam}}(R), 1\}} \max_{\substack{w \in Q^\top(C-x^0) \\ \|w\| \le \min\{\overline{\operatorname{diam}}(R),1\}}} \|Mw\|$$

where $\overline{\operatorname{diam}}(R) := \max_{r_i \in R} \|r_i\|$

Note: If $C = \mathbb{R}^n$, this reduces to the spectral norm: $\|M\|_{x^0,C,D} = \|M\|$

A geometry measure that respects the feasible set

Let D consist of sample directions and write $D = QR$, where $Q^\top Q = I_p$

Feasible geometry measure: matrix spectral norm w.r.t. $Q^\top(C - x^0)$

For any matrix $M \in \mathbb{R}^{m\times p}$, define

$$\|M\|_{x^0,C,D} := \frac{1}{\min\{\overline{\operatorname{diam}}(R), 1\}} \max_{\substack{w\in Q^\top(C-x^0)\\ \|w\|\le \min\{\overline{\operatorname{diam}}(R),1\}}} \|Mw\|$$

where $\overline{\operatorname{diam}}(R) := \max_{r_i\in R} \|r_i\|$

Note: If $C = \mathbb{R}^n$, this reduces to the spectral norm: $\|M\|_{x^0,C,D} = \|M\|$

Why this fixes the skinny-feasible-set issue

For $C = \{(x_1, x_2) : |x_2| \le \epsilon\}$, $x^0 = \mathbb{0}$, $D = I_2$, and $M = \operatorname{Diag}(1, \epsilon^{-1})$,

$$\|M\|_{\mathbb{0},C,I_2} = \max_{|w_2|\le\epsilon, \|w\|\le 1} \sqrt{w_1^2 + \epsilon^{-2} w_2^2} = \sqrt{2 - \epsilon^2} \le \sqrt{2}$$

Constructing (C, Q)-fully linear models

Let D consist of sample directions and write $D = QR$, where $Q^\top Q = I_p$

Theorem

If $f \in \mathcal{C}^{1+}$, D has full column rank, and

$$\left\|\overline{R}^{-1}\right\|_{x^0,C,D} \quad \textit{is uniformly bounded, where } \overline{R} := R/\overline{\operatorname{diam}}(R),$$

then the subspace linear interpolation model $\widehat{m}$ is (C, Q)-fully linear in $B(x^0; \overline{\operatorname{diam}}(R))$

What does stationarity mean with constraints?

Definition (Conn, Gould, Toint, 2000)

First-order criticality measure for convex-constrained optimization

$$\pi^f(x) := \left| \min_{\substack{x+d \in C \\ \|d\| \leq 1}} \nabla f(x)^\top d \right|$$

Interpretation

$\pi^f(x)$ measures the maximum first-order feasible decrease at x

Random subspaces must preserve feasible descent

Let $x + \mathrm{col}(D)$ be the affine subspace and decompose $D = QR$

Definition (Chen, Hare, Wiebe, 2025)

Let $\alpha \in (0, 1)$. We say that $D \in \mathbb{R}^{n \times p}$ is α-well-aligned for f and C at x if

$$\left| \min_{\substack{d \in C - x \\ \|d\| \leq 1}} \nabla f(x)^\top Q Q^\top d \right| \geq \alpha \pi^f(x)$$

Random subspaces must preserve feasible descent

Let $x + \mathrm{col}(D)$ be the affine subspace and decompose $D = QR$

Definition (Chen, Hare, Wiebe, 2025)

Let $\alpha \in (0, 1)$. We say that $D \in \mathbb{R}^{n \times p}$ is α-well-aligned for f and C at x if

$$\left| \min_{\substack{d \in C - x \\ \|d\| \leq 1}} \nabla f(x)^\top Q Q^\top d \right| \geq \alpha \pi^f(x)$$

Theorem (Chen, Hare, Wiebe, 2025)

(Idea: Concentration of measure on the Grassmannian)

Suppose $\beta \in (0, 1)$, $p \in (0, n]$, *and i.i.d.* $D_{ij} \sim \mathcal{N}(0, 1)$. *Then,*

$$\mathbb{P}\left[D \text{ is } \left(\frac{p}{n}\beta\right)\text{-well-aligned for } f \text{ and } C \text{ at } x\right]$$

$$\geq \text{complicated stuff that depends on } n, p, \beta, \pi^f(x), \text{ and } \|\nabla f(x)\|$$

Convergence and complexity results

Let $\epsilon > 0$, (UC)=UnConstrained, and (CC)=Convex-Constrained

- (UC) $\mathbb{P}\left[\min_{k \le K} \|\nabla f(x_k)\| < \epsilon\right] \ge 1 - e^{-C(K+1)}$

  (CC) $\mathbb{P}\left[\min_{k \le K} \pi^f(x_k) < \epsilon\right] \ge 1 - e^{-C(\epsilon)(K+1)}$

- (UC) $\mathbb{P}\left[\inf_{k \ge 0} \|\nabla f(x_k)\| = 0\right] = 1$

  (CC) $\mathbb{P}\left[\inf_{k \ge 0} \pi^f(x_k) = 0\right] = 1$

- (UC) $\mathbb{E}\left[\min\left\{k \ge 0 : \|\nabla f(x_k)\| < \epsilon\right\}\right] = \mathcal{O}(\epsilon^{-2})$

  (CC) $\mathbb{E}\left[\min\left\{k \ge 0 : \pi^f(x_k) < \epsilon\right\}\right] = \mathcal{O}(\epsilon^{-4})$

(UC) from [Cartis, Roberts, 2023]; (CC) from [Chen, Hare, Wiebe, 2025]

Comparison of runtime to reach f^*

Steps:

1. Run CLARSTA for $100(n+1)$ fevals and denote the result by f^*
2. Run COBYLA until f^* is reached or $100(n+1)$ fevals or 10^5 seconds are required

Results when $n = 1000$:

- In some problems, COBYLA uses less fevals to reach f^*
- COBYLA requires more than 1000 times more time than CLARSTA to reach f^* (1min vs. 17h)

Results when $n = 10000$:

- CLARSTA reaches f^* in 30 minutes
- COBYLA never reaches f^* in 10^5 seconds ($\sim$ 28h)

Outline

Summary

In summary,

- High-dimensional DFO problems can be addressed efficiently by random subspace methods
- Subspace modeling provides a bridge between practical scalability and convergence theory
- QARSTA is a random subspace MBTR method for unconstrained problems, using low-dimensional quadratic models
- CLARSTA extends the unconstrained framework to convex constraints

Possible future directions:

- Quadratic models in CLARSTA
- Better complexity bounds
- Nonconvex/blackbox constraints

Thank you

- Yiwen Chen. *Relationships between full-space and subspace quadratic interpolation models and simplex derivatives*. 2026. arXiv: 2602.10374
- Yiwen Chen, Warren Hare, and Amy Wiebe. "Q-fully quadratic modeling and its application in a random subspace derivative-free method". In: *Computational Optimization and Applications* 89.2 (2024), pp. 317–360
- Yiwen Chen, Warren Hare, and Amy Wiebe. *CLARSTA: A random subspace trust-region algorithm for convex-constrained derivative-free optimization*. 2025. arXiv: 2506.20335

Email: `yiwchen@student.ubc.ca`