Randomized subspace methods for high-dimensional model-based derivative-free optimization

Yiwen Chen

Department of Mathematics
University of British Columbia

AustMS 2025, Melbourne, Australia

December 10, 2025

Based on joint works with Warren Hare and Amy Wiebe

Outline

1. Introduction
2. Random subspace model-based trust-region algorithm
3. Constraints?
4. Summary

Derivative-free optimization (DFO)

Consider the optimization problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

where f is given by a blackbox:

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

Note: It does not mean that the derivatives do not exist

Model-based DFO

Model-based DFO methods:

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

Limitations:

- Number of function evals. is too high for large problems ($n \approx 1000$)

n	1	10	100	1000
$(n+1)(n+2)/2$	3	66	5151	501501

- Primarily designed for small- to medium-scale problems ($n \leq 100$)

Randomized subspace model-based DFO

Idea:

1. Select a low-dimensional affine subspace
2. Build and optimize a model to compute a step in this subspace
3. Change the affine subspace at the next iteration

Some existing papers:
[Zhang, 2012]; [Cartis, Roberts, 2023]; [Dzahini, Wild, 2024];
[Chen, Hare, Wiebe, 2024]; [Cartis, Roberts, 2024]
[Chen, Hare, Wiebe, 2025]

Outline

Model-based trust-region (MBTR) algorithm

for $k = 0, 1, \dots$ **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

$$\rho^k = \frac{f(x^k) - f(x^k + s^k)}{m^k(\mathbb{0}) - m^k(s^k)} = \frac{true \ \ decrease}{predicted \ \ decrease}$$

Accept/reject step based on ρ^k and update trust region radius

Random subspace MBTR algorithm

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

Define an affine subspace $x^k + D^k\mathbb{R}^p$ by selecting $D^k \in \mathbb{R}^{n\times p}$
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

Model construction

Q-fully quadratic models

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 $\|s\| \leq \Delta$,

$$
\begin{aligned}
|f(x+s) - m(x+s)| &\leq \kappa_f \Delta^3 \\
\|\nabla f(x+s) - \nabla m(x+s)\| &\leq \kappa_g \Delta^2 \\
\left\|\nabla^2 f(x+s) - \nabla^2 m(x+s)\right\| &\leq \kappa_h \Delta
\end{aligned}
$$

Q-fully quadratic models

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 quadratic models

Definition. [Custódio, Dennis Jr., Vicente, 2008] & [Hare, Jarry-Bolduc, Planiden, 2023]
Let $x^0 \in \mathbb{R}^n$ and $D = [d^1 \cdots d^p] \in \mathbb{R}^{n \times p}$
The generalized simplex gradient of f at x^0 over D is defined by

$$\nabla_s f(x^0; D) = \left(D^\top\right)^\dagger \delta_f(x^0; D)$$

where

$$\delta_f(x^0; D) = \begin{bmatrix} f(x^0 + d^1) - f(x^0) \\ f(x^0 + d^2) - f(x^0) \\ \vdots \\ f(x^0 + d^p) - f(x^0) \end{bmatrix}$$

The generalized simplex Hessian of f at x^0 over D is defined by

$$\nabla_s^2 f(x^0; D) = \left(D^\top\right)^\dagger \delta_{\nabla_s f}(x^0; D)$$

Constructing Q-fully quadratic models

Theorem. [Chen, Hare, Wiebe, 2024]
Let $x^0 \in \mathbb{R}^n$ and $D = QR$ have full col. rank, where $R = [r^1 \cdots r^p] \in \mathbb{R}^{p \times p}$
Let

$$m(x) = f(x^0) + \left(2\nabla_s f(x^0; D) - \nabla_s f(x^0; 2D)\right)^\top \left(x - x^0\right) + \frac{1}{2}\left(x - x^0\right)^\top \nabla_s^2 f(x^0; D; D)\left(x - x^0\right)$$

Then, the model $\widehat{m} : \mathbb{R}^p \to \mathbb{R}$ defined by $\widehat{m}(\widehat{s}) = m(x^0 + Q\widehat{s})$ is Q-fully quadratic in $B(x^0, \max_{1 \le i \le p} \|r^i\|)$

The $2\nabla_s f(x^0; D) - \nabla_s f(x^0; 2D)$ is a special case of the *Adapted Centred Simplex Gradient*, see Y. Chen and W. Hare. "Adapting the centred simplex gradient to compensate for misaligned sample points". In: *IMA J. Numer. Anal.* (2023)

Subspace selection

α-well-aligned matrices

Let $x + D\mathbb{R}^p$ 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

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

α-well-aligned matrices

Let $x + D\mathbb{R}^p$ 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)
Let $\alpha, \delta \in (0, 1)$. Suppose $p \geq 4(1 - \alpha)^{-2} \ln(1/\delta)$ and let $D_{ij} \sim \mathcal{N}(0, 1/p)$
Then,

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

Can we reuse past information to construct subspaces?

Suppose D has the form $D = [D^U \; D^R] \in \mathbb{R}^{n \times p}$, where

- $D^U \in \mathbb{R}^{n \times (p - p_{\text{rand}})}$ is picked from previous sample points
- $D^R \in \mathbb{R}^{n \times p_{\text{rand}}}$ is randomly generated

Theorem. [Chen, Hare, Wiebe, 2024]
Let $\alpha, \delta \in (0, 1)$
Suppose $p_{\text{rand}} \geq 4(1 - \alpha)^{-2} \ln(1/\delta)$ and let

$$D^R \sim \operatorname{proj}_{\operatorname{col}(D^U)^\perp} \left(\mathcal{MN} \left(\mathbb{0}, \frac{1}{p} I_n, I_{p_{\text{rand}}} \right) \right)$$

Then, there exists $\alpha_D > 0$ such that

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

Outline

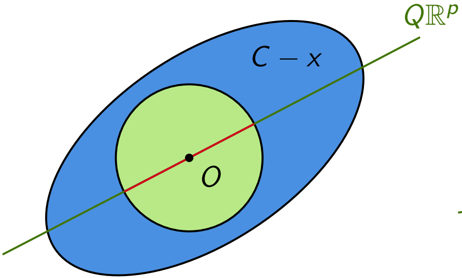
(C, Q)-fully linear models

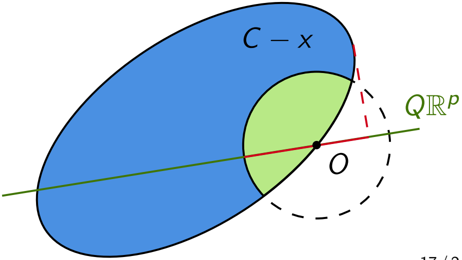
Let C be the constraint set (convex, closed, nonempty interior)

Definition. [Chen, Hare, Wiebe, 2025]
Let $Q \in \mathbb{R}^{n \times p}$. A model $\widehat{m} : \mathbb{R}^p \to \mathbb{R}$ is (C, Q)-fully linear in $B(x, \Delta)$ if there exist $\kappa_f, \kappa_g > 0$ s.t. for all $\widehat{s} \in Q^\top (C - x)$ with $\|\widehat{s}\| \leq \Delta$,

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

$$\max_{\substack{d \in Q^\top (C - x) \\ \|d\| \leq 1}} \left| \left(Q^\top \nabla f(x + Q\widehat{s}) - \nabla \widehat{m}(\widehat{s}) \right)^\top d \right| \leq \kappa_g \Delta$$





Constructing (C, Q)-fully linear models

Theorem. [Chen, Hare, Wiebe, 2025]
Let $x^0 \in \mathbb{R}^n$ and $D = QR$ have full col. rank, where $R = [r^1 \cdots r^p] \in \mathbb{R}^{p \times p}$
Let

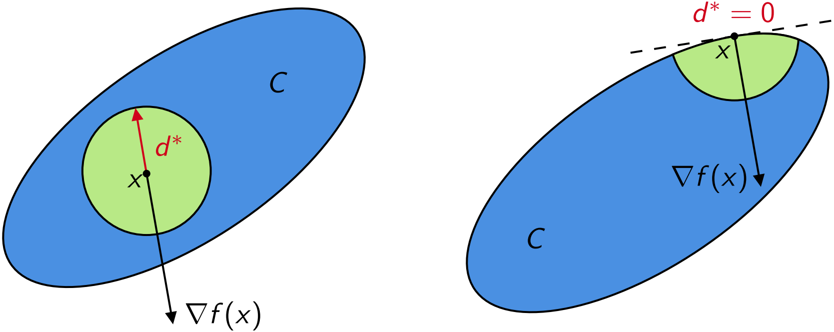
$$m(x) = f(x^0) + \nabla_s f(x^0; D)^\top \left(x - x^0\right)$$

Then, the model $\widehat{m} : \mathbb{R}^p \to \mathbb{R}$ defined by $\widehat{m}(\widehat{s}) = m(x^0 + Q\widehat{s})$ is (C, Q)-fully linear in $B(x^0, \max_{1 \le i \le p} \|r^i\|)$

First-order criticality measure

First-order criticality measure for convex-constrained optimization
[Conn, Gould, Toint, 2000]

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



α-well-aligned matrices (convex-constrained version)

Let $x + D\mathbb{R}^p$ be the affine subspace and $D = QR$ be the QR factorization

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

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

α-well-aligned matrices (convex-constrained version)

Let $x + D\mathbb{R}^p$ be the affine subspace and $D = QR$ be the QR factorization

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

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

Theorem. [Chen, Hare, Wiebe, 2025]
(Idea: Concentration on the Grassmannian)
Suppose $p \geq n\alpha$ and let $D_{ij} \sim \mathcal{N}(0, 1)$. Then,

$$\mathbb{P}\left[D \text{ is } \alpha\text{-well-aligned for } f \text{ and } C \text{ at x}\right]$$
$$\geq \text{complicated stuff that depends on } n,\ p,\ \alpha,\ \pi^f(x), \text{ and } \|\nabla f(x)\|$$

Convergence and complexity results

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

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

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

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

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

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

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

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

Outline

Summary

In summary,

- High-dimensional DFO problems are hard
- Unconstrained problems can be effectively approached by randomized subspace methods
- Randomized subspace methods work for convex constrained problems, but the projection onto the feasible set is required

Future directions:

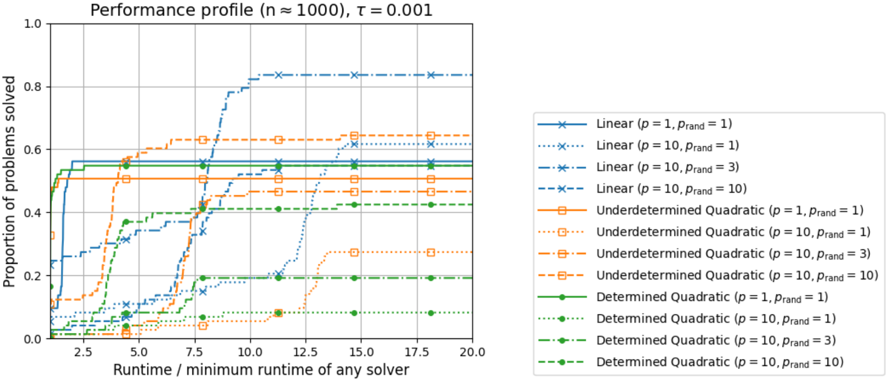
- Nonconvex constraints?
- Blackbox constraints?
- Random manifolds?

Thank you

- Y. Chen, W. Hare, and A. 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
- Y. Chen, W. Hare, and A. Wiebe. "CLARSTA: A random subspace trust-region algorithm for convex-constrained derivative-free optimization". In: *arXiv preprint arXiv:2506.20335* (2025)

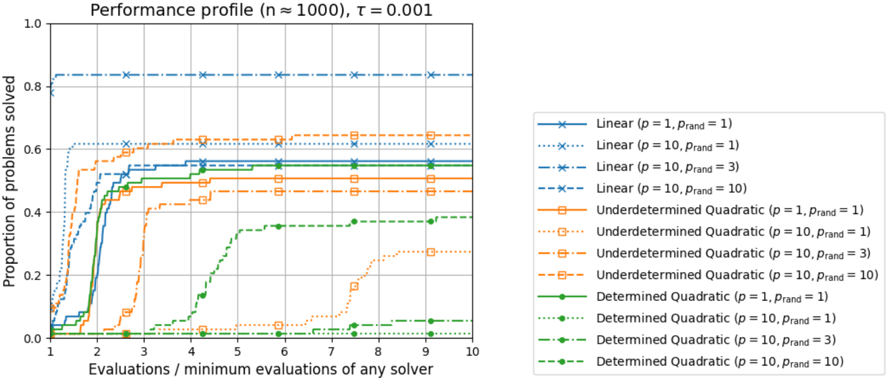
Email: `yiwchen@student.ubc.ca`

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 fevals



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

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 $n = 1000$:

Problem		CLARSTA		COBYLA	
obj.	const.	nf	Total time (s)	nf	Total time (s)
C.R.	box	100100	6.451e+01	26716	6.196e+04
C.R.	ball	100100	6.913e+01	26337	2.048e+04
C.R.	halfspace	100100	7.349e+01	27054	2.113e+04
Trig.	box	100100	6.975e+03	N/A	1.000e+05*
Trig.	ball	100100	7.052e+03	100100*	8.678e+04
Trig.	halfspace	100100	6.901e+03	100100*	8.655e+04

*COBYLA does not reach f^*

Comparison of runtime to reach f^*

Results when $n = 10000$:

Problem		CLARSTA		COBYLA	
obj.	const.	nf	Total time (s)	nf	Total time (s)
C.R.	box	1000100	1.836e+03	N/A	1.000e+05*
C.R.	ball	1000100	1.960e+03	N/A	1.000e+05*
C.R.	halfspace	1000100	1.928e+03	N/A	1.000e+05*

*COBYLA does not reach f^*