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

Yiwen Chen

Department of Mathematics
University of British Columbia (Okanagan)

WCOM 2025, Kelowna, Canada

October 18, 2025

Based on joint works with Warren Hare and Amy Wiebe

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 ($n \approx 1000$)
- 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

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

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

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

Model quality requirement: Q-fully linear models

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

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

Model quality requirement: Q-fully linear models

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

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

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 $\kappa_f(x), \kappa_g(x) > 0$ s.t. for all $\widehat{s} \in \mathbb{R}^p$ with $\|\widehat{s}\| \leq \Delta$,

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

Note: Check out [Chen, Hare, Wiebe, 2024] for construction & management procedures!

Subspace quality requirement: 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

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

Subspace quality requirement: 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,

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

A quick comparison

	[CartisRoberts23]	[DzahiniWild24]	[ChenHareWiebe24]
Problem	Deterministic	Stochastic	Deterministic
Model	Linear	Linear	**Quadratic**
Sample set	Completely resample	Completely resample	**Reuse previous points**
Subspace	No discussion on p	Lower bound on p	Lower bound on p_{rand}

Looks good, but...

Haha, my problem has constraints!

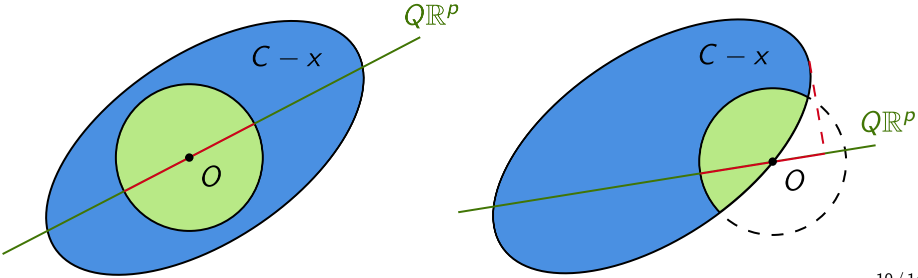
(C, Q)-fully linear models

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 $\kappa_f(x), \kappa_g(x) > 0$ s.t. for all $\widehat{s} \in Q^\top(C - x)$ with $\|\widehat{s}\| \leq \Delta$,

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

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



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

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

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

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

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

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

Convergence and complexity results

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

- (UC) $\mathbb{P}\left[\inf_{k\geq 0} \|\nabla f(x_k)\| = 0\right] = 1$
  (CC) $\mathbb{P}\left[\inf_{k\geq 0} \pi^f(x_k) = 0\right] = 1$
- (UC) $\mathbb{E}\left[\min\left\{k \geq 0 : \|\nabla f(x_k)\| < \epsilon\right\}\right] = \mathcal{O}(\epsilon^{-2})$
  (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]

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`