CLARSTA: A Random Subspace Trust-region Algorithm for Convex-constrained Derivative-free Optimization

Yiwen Chen

Department of Mathematics
University of British Columbia

DFOS26, Montréal, Canada

June 22, 2026

Based on joint work with Warren Hare and Amy Wiebe
Includes joint work with Lindon Roberts

High-dimensional model-based DFO

We consider

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

where $\Omega \subseteq \mathbb{R}^n$ and f is given by a black box:

Model-based DFO methods:

- Construct surrogate models to approximate objective function

Issue in high dimensions: full-space model cost $\sim \mathcal{O}(n)$ or $\mathcal{O}(n^2)$

Idea:

$$\mathbb{R}^n \quad \rightsquigarrow \quad x_k + \mathrm{col}(Q_k) \;\text{ where }\; Q_k \in \mathbb{R}^{n \times p},\; Q_k^\top Q_k = I_p,\; p \ll n$$

CLARSTA at a glance

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 $\mathbb{R}^p$:

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

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

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

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

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

General framework of CLARSTA

for $k = 0, 1, \dots$ **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 $\mathbb{R}^p$:

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

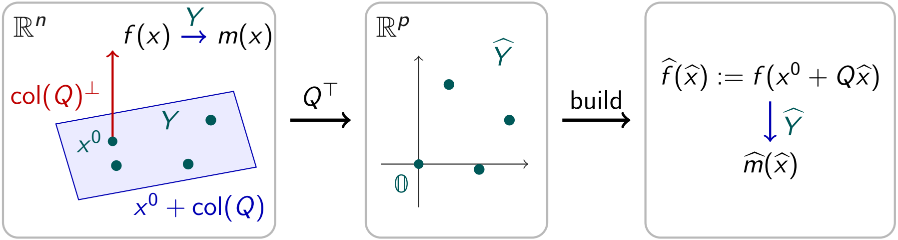
and calculate the corresponding feasible step $s_k \in \mathbb{R}^n$

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

Subspace vs. full-space models



Lift and equivalence (Chen 2026)

For $\mathcal{X} \in \{\mathrm{MN}, \mathrm{MFN}, \ldots\}$, we have

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

and

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

From fully linear to (C, Q)-fully linear

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*

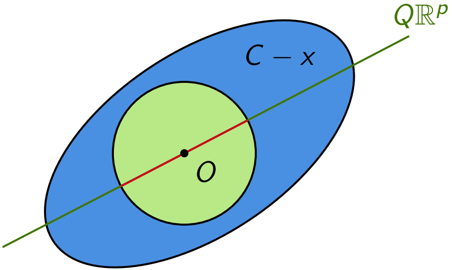
(C, Q)-fully linear models

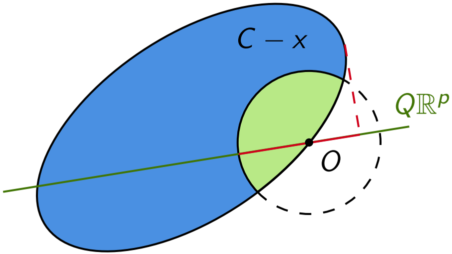
Definition (Chen-Hare-Wiebe 2025)

Let $Q \in \mathbb{R}^{n \times p}$ consist of p orthonormal columns
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 $\|\widehat{s}\| \le \Delta$,

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

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





Classical geometry measures can blow up

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)

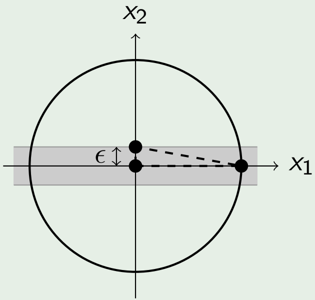
Consider

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

and sample set

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

Classical geometry measures are $\sim \epsilon^{-1}$



Feasible sample set geometry measure

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

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

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

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

Feasible sample set geometry measure

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

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

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

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 $M = \mathrm{Diag}(1, \epsilon^{-1})$,

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

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

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

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

α-well-aligned matrices (unconstrained version)

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,*

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

First-order criticality measure

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

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

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

  (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 $\mathrm{CLARSTA}$ 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 $\mathrm{CLARSTA}$ to reach f^* (1min vs. 17h)

Results when $n = 10000$:

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

Toward quadratic CLARSTA

Quadratic models

Minimum norm (MN), minimum Frobenius norm (MFN), quadratic generalized simplex derivative (QS) models, etc.

Fully linear bounds

MN and MFN bounds were known previously
In [Chen-Hare-Roberts 2026], we

- remove the bounded model-Hessian assumption for MN
- give the first fully linear bounds for QS

Directional Hessian accuracy

With a mild sample set structure, all three models capture fully quadratic accuracy along sample directions

Directional Hessian accuracy

Hessian error ranges from $\mathcal{O}(1)$ to $\mathcal{O}(\Delta_Y)$ (Chen-Hare-Roberts 2026)

Suppose that $f \in \mathcal{C}^{2+}$,

$$Y = \left\{x^0, x^0 \pm d_i : i \in \{1, \ldots, p\}\right\}, \quad \Delta_Y := \max_i \|d_i\|, \quad \overline{d} := d/\|d\|$$

For $\mathcal{X} \in \{\mathrm{MN}, \mathrm{MFN}, \mathrm{QS}\}$,

$$\left|\overline{d_i}^\top \left(\nabla^2_{\mathcal{X}} f(x^0; Y) - \nabla^2 f(x^0)\right) \overline{d_i}\right| = \mathcal{O}(\Delta_Y)$$

Directional Hessian accuracy

Hessian error ranges from $\mathcal{O}(1)$ to $\mathcal{O}(\Delta_Y)$ (Chen-Hare-Roberts 2026)

Suppose that $f \in \mathcal{C}^{2+}$,

$$Y = \left\{x^0, x^0 \pm d_i : i \in \{1, \dots, p\}\right\}, \quad \Delta_Y := \max_i \|d_i\|, \quad \overline{d} := d/\|d\|$$

For $\mathcal{X} \in \{\mathrm{MN}, \mathrm{MFN}, \mathrm{QS}\}$,

$$\left|\overline{d_i}^\top \left(\nabla^2_{\mathcal{X}} f(x^0; Y) - \nabla^2 f(x^0)\right) \overline{d_i}\right| = \mathcal{O}(\Delta_Y)$$

In general, for any non-zero $d \in \mathrm{col}(D)$, and $v := D^\dagger d$,

$$\begin{aligned}&\left|\overline{d}^\top \left(\nabla^2_{\mathcal{X}} f(x^0; Y) - \nabla^2 f(x^0)\right) \overline{d}\right| \\ &= \mathcal{O}\left(\frac{\|v\|_1^2 - \|v\|_\infty^2}{\|v\|^2}\right) + \mathcal{O}\left(\frac{\|v\|_1^2 - \|v\|_\infty^2}{\|v\|^2}\Delta_Y\right) + \mathcal{O}(\Delta_Y)\end{aligned}$$

Summary

CLARSTA is a new approach to large-scale convex-constrained DFO:

- Extends unconstrained framework using feasible subspace geometry and (C, Q)-fully linear models
- Samples random subspaces that are sufficiently aligned with feasible descent directions, yielding convergence and complexity bounds
- Demonstrates competitive high-dimensional performance on problems with up to 10000 variables

Ongoing: incorporate quadratic models into CLARSTA :)

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 Lindon Roberts. *Accuracy and Relationships of Quadratic Models in Derivative-free Optimization*. 2026. arXiv: 2605.12819
- 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