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

Yiwen Chen

Department of Mathematics
University of British Columbia

IOS 2026, Atlanta, United States

March 20, 2026

Joint work with Warren Hare and Amy Wiebe

Model-based derivative-free optimization (DFO)

We consider

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

where C is convex and closed, and f is given by a black box:

Model-based DFO methods:

- Construct surrogate models to approximate objective function

Limitations:

- Number of function evals. is too high for large problems ($n >> 100$)
- Primarily designed for small- to medium-scale problems ($n \leq 100$)

Randomized subspace model-based DFO

Key idea:

1. Sample a p-dimensional affine subspace
2. Build model in the subspace
3. Solve trust-region subproblem in the subspace

Some related works:

[Zhang, 2012]; [Cartis, Roberts, 2023]; [Dzahini, Wild, 2024];
[Chen, Hare, Wiebe, 2024]; [Cartis, Roberts, 2024];
...

General framework of CLARSTA

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 $\mathbb{R}^p$:

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

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

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

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

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

Model construction

Fully linear models

Standard trust-region DFO requires fully linear models:
For all $\|s\| \leq \Delta$,

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

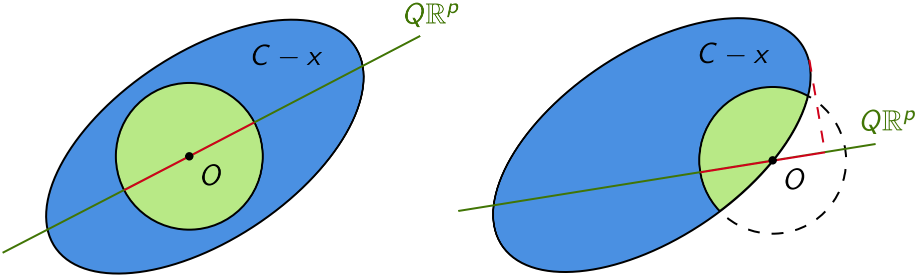
Problem: [Larson, Menickelly, Wild, 2019] "It may be impossible to obtain a fully linear model using only feasible points"

(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 $\|\widehat{s}\| \leq \Delta$,

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

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



Constructing (C, Q)-fully linear models

Models are built using linear interpolation

Steps:

1. Sample full-column rank $D \in \mathbb{R}^{n \times p}$
2. Compute generalized simplex gradient $\nabla_s f(x_k; D)$ and construct

$$m(x) = f(x_k) + \nabla_s f(x_k; D)^\top (x - x_k)$$

3. Decompose $D = QR$ and restrict $m(x)$ to subspace

$$\widehat{m}(\widehat{s}) = m(x_k + Q\widehat{s})$$

The model $\widehat{m}(\widehat{s})$ is proven to be (C, Q)-fully linear

---

$\widehat{m}(\widehat{s})$ is the determined linear interpolation model of $\widehat{f}(\widehat{s}) = f(x_k + Q\widehat{s})$ over R

Subspace selection

α-well-aligned matrices (unconstrained version)

Random subspace must preserve criticality

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

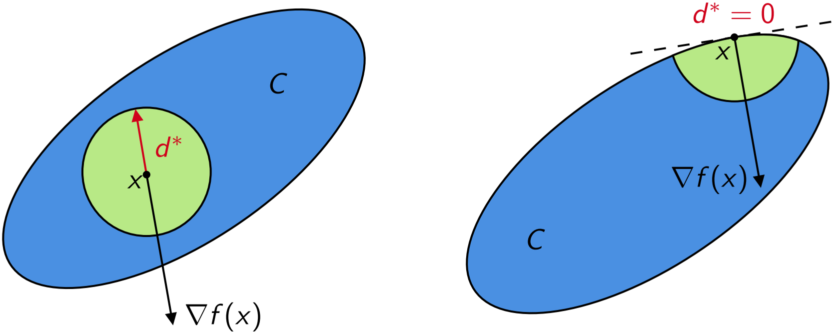
Definition. [Cartis, Roberts, 2023]
Let $\alpha \in (0, 1)$. We say that $D \in \mathbb{R}^{n \times p}$ is α-well-aligned for f at x if

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

First-order criticality measure

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

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



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

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

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

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

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

  (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]

Numerical results

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^*

Summary

CLARSTA provides a new approach to large-scale constrained DFO:

- Handles general convex constraints within a trust-region framework
- Almost-sure global convergence and complexity guarantees
- Competitive performance in high dimensions, demonstrated up to 10000 variables

Future directions:

- Improve complexity result from $\mathcal{O}(\epsilon^{-4})$ to $\mathcal{O}(\epsilon^{-2})$
- Nonconvex/blackbox constraints?
- Random manifolds?

Thank you

- 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