Random subspace trust-region algorithm for high-dimensional convex-constrained derivative-free optimization

Yiwen Chen

Department of Mathematics
University of British Columbia

June, 2025

Joint work with Dr. Warren Hare and Dr. Amy Wiebe

Outline

Derivative-free optimization (DFO)

Consider the optimization problem

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

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:

- Function evaluations are too expensive for large problems ($n \approx 1000$)
- Primarily designed for small- to medium-scale problems ($n \leq 100$)

Random subspace model-based DFO (unconstrained)

Idea:

1. Select a low-dimensional affine subspace
2. Build and minimize 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]

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 $\mathbb{R}^n$:

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

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

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

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

Fully linear models

Definition. A model $m : \mathbb{R}^n \to \mathbb{R}$ is fully linear in $B(x, \Delta) \subseteq \mathbb{R}^n$ if there exist constants $\kappa_{ef}(x) > 0$ and $\kappa_{eg}(x) > 0$ such that for all $s \in \mathbb{R}^n$ with $\|s\| \leq \Delta$,

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

Constructing fully linear models

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

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

$$\|\nabla f(x+s) - \nabla f(x)\| \leq L_{\nabla f} \|s\|$$

Constructing fully linear models

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

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

Construct a fully linear model $m(x+s) = f(x) + g^\top s$ where $g \approx \nabla f(x)$

- linear interpolation
- linear regression
- ...

Outline

Convex-constrained 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 $\mathbb{R}^p$:

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

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

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

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

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

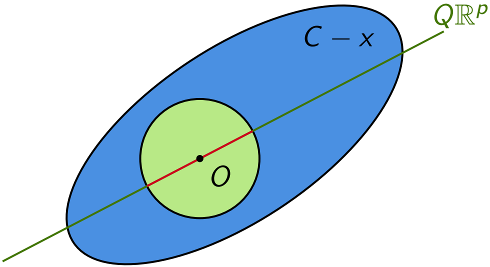
Model construction

(C, Q)-fully linear models

Definition. 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 $B(x, \Delta) \subseteq \mathbb{R}^n$ if there exist constants $\kappa_{ef}(x) > 0$ and $\kappa_{eg}(x) > 0$ such that for all $\widehat{s} \in Q^\top(C - x)$ with $\|\widehat{s}\| \leq \Delta$,

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

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



A generalization of fully linear models

Note: If $C = \mathbb{R}^n$ and $Q = I_n$, then for all $\|\widehat{s}\| \leq \Delta$

$$|f(x+\widehat{s}) - \widehat{m}(\widehat{s})| \leq \kappa_{ef}(x)\Delta^2$$
$$\|\nabla f(x+\widehat{s}) - \nabla\widehat{m}(\widehat{s})\| \leq \kappa_{eg}(x)\Delta$$

which aligns with the definition of fully linear models

Actually, this is also a generalization of Q-fully linear models and C-pointwise fully linear models (see details in manuscript)

Constructing (C, Q)-fully linear models

Let $f \in \mathcal{C}^{1+}$
Let $x + D\mathbb{R}^p$ be the affine subspace and $D = QR$ be the QR factorization
Denote $R = [r_1 \cdots r_p]$ and $\overline{\mathrm{diam}}(R) = \max_{1 \leq i \leq p} \|r_i\|$

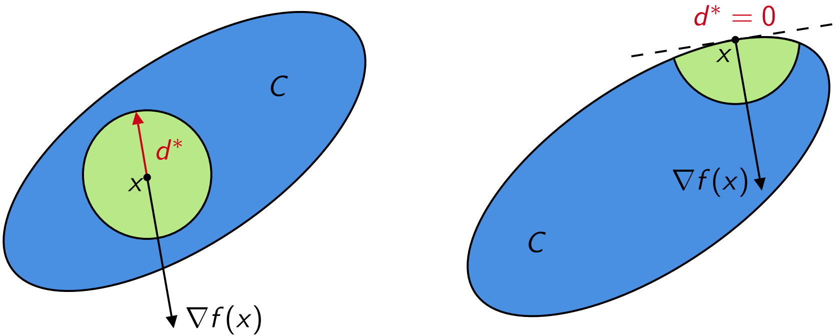
Theorem. Let $\widehat{f}(\widehat{s}) = f(x + Q\widehat{s})$ and $\widehat{m} : \mathbb{R}^p \to \mathbb{R}$ be the *determined linear interpolation* model of $\widehat{f}$ on $\{\mathbb{0}_p\} \cup \{r_i : i = 1, \ldots, p\}$
Then, $\widehat{m}$ is (C, Q)-fully linear in $B(x, \overline{\mathrm{diam}}(R))$

Subspace selection

First-order criticality measure

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

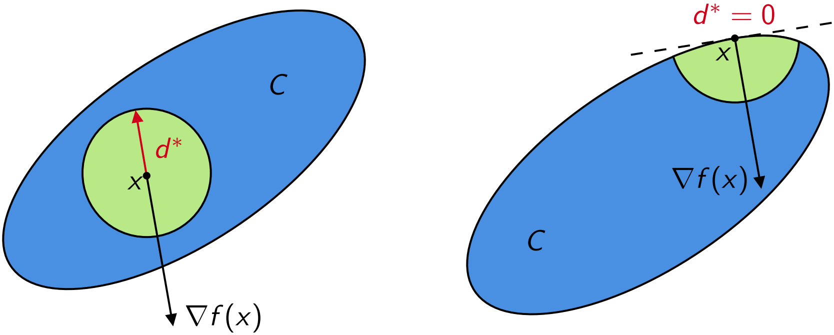
$$\pi^f(x) = \left| \min_{\substack{x+d\in C \\ \|d\|\leq 1}} \nabla f(x)^\top d \right|$$



First-order criticality measure

First-order criticality measure for convex-constrained optimization

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



α-well-aligned matrices

Let $f \in \mathcal{C}^1$ and $\alpha \in (0,1)$
Let $x + D\mathbb{R}^p$ be the affine subspace and $D = QR$ be the QR factorization

Unconstrained version: [Cartis, Roberts, 2023]
Definition. 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 $f \in \mathcal{C}^1$ and $\alpha \in (0, 1)$
Let $x + D\mathbb{R}^p$ be the affine subspace and $D = QR$ be the QR factorization

Unconstrained version: [Cartis, Roberts, 2023]
Definition. We say that $D \in \mathbb{R}^{n \times p}$ is α-well-aligned for f at x if

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

Convex-constrained version:
Definition. We say that $D \in \mathbb{R}^{n \times p}$ is α-well-aligned for f and C at x if

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

Constructing α-well-aligned matrices

Let $f \in \mathcal{C}^1, x \in \mathbb{R}^n$, and $\alpha \in (0,1)$

Theorem. Suppose $p \geq n\alpha$. Let $D_{ij} \sim \mathcal{N}(0,1)$. Then,

$$\mathbb{P}\left[D \text{ is } \alpha\text{-well-aligned for } f \text{ and } C \text{ at x}\right]$$
$$\geq \begin{cases} 1, & \text{if } \pi^f(x) = 0, \\ 1 - \exp\left(-\frac{n-1}{8}\left(\frac{p}{n} - \alpha\right)^2 \left(\frac{\pi^f(x)}{\|\nabla f(x)\|}\right)^2\right), & \text{if } \pi^f(x) \neq 0 \end{cases}$$

Convergence

Assumptions

- $f \in \mathcal{C}^{1+}$ and has compact feasible level set $C \cap \mathrm{lev}_{f(x_0)}$
- $\|\nabla^2 \widehat{m}_k\| \leq \kappa_H$ for all k
- $\exists \kappa_{\mathrm{tr}} \in (0,1)$ s.t. the solution $\widehat{s}_k$ of the trust-region subproblem satisfy

$$\widehat{m}_k(\mathbb{0}_p) - \widehat{m}_k(\widehat{s}_k) \geq \kappa_{\mathrm{tr}} \pi^m(x_k) \min\left(\frac{\pi^m(x_k)}{\|\nabla^2 \widehat{m}_k\| + 1}, \Delta_k, 1 \right)$$

- $\exists \overline{k} \geq 0$ s.t. the trust-region radius are not increased for all $k \geq \overline{k}$

Convergence results

Theorem. For all $\epsilon > 0$, there exist $C > 0$ and a sufficiently large K s.t.

$$\mathbb{P}\left[\min_{k \leq K} \pi^f(x_k) < \epsilon\right] \geq 1 - e^{-C(K-\overline{k}+1)\epsilon^2}$$

Theorem. If $\mathrm{CLARSTA}$ is run with $\Delta_{\min} = 0$, then

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

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

The $\mathbb{P}[\cdot]$ and $\mathbb{E}[\cdot]$ give the probability and expected value of a random variable

Outline

Test settings

Solvers:

- CLARSTA
- COBYLA

Test functions:

- ChainRosenbrock: $f(x) = \sum_{i=1}^{n-1} (100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2)$
- Trigonometric: $f(x) = \sum_{i=1}^{n} (n - \sum_{j=1}^{n} \cos x_j + i(1 - \cos x_i) - \sin x_i)^2$

Constraints:

- box with x^* at the corner
- ball with x^* on the boundary
- half-space with x^* in the interior

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^*

Relative performance vs. per-feval time

Suppose that

$$\mathrm{TotalTime}_{\mathrm{alg}} = \mathrm{AlgTime}_{\mathrm{alg}} + \mathrm{nf}_{\mathrm{alg}} \mathrm{t}_{\mathrm{feval}}$$

Then, we can compute the maximum per-feval time for $\mathrm{CLARSTA}$ to be faster than COBYLA

Results (in seconds):

obj.	const.	$n = 100$	$n = 300$	$n = 900$	$n = 1000$
C.R.	box	9.605e-04	3.447e-02	1.921e+00	8.435e-01
C.R.	ball	0	8.727e-03	6.324e-01	2.769e-01
C.R.	halfspace	0	7.775e-03	6.182e-01	2.884e-01
Trig.	box	0	1.339e-01	1.375e+03	+Inf
Trig.	ball	0	5.226e-02	+Inf	+Inf
Trig.	halfspace	0	3.057e-02	+Inf	+Inf

Outline

Summary

For high-dimensional convex-constrained DFO problems, we:

- Defined a class of (C, Q)-fully linear models that is easy to analyze and construct
- Provided a new subspace sampling technique that preserves first-order criticality measure by a certain percentage
- Proposed an algorithm with convergence analysis and reliable performance in high dimensions

Future directions:

- Reduce worst-case complexity
- "(C, Q)-fully quadratic"

Thank you

- Yiwen Chen, Warren Hare, Amy Wiebe. *CLARSTA: A random subspace trust-region algorithm for convex-constrained derivative-free optimization.* (Preprint available soon.)
  Email: `yiwchen@student.ubc.ca`