Subspace Modeling and Random Subspace Trust-region Methods in Derivative-free Optimization

Yiwen Chen

Department of Mathematics
University of British Columbia

Optimization 2026, Lisbon, Portugal

July 21, 2026

Joint works with Warren Hare and Amy Wiebe

Outline

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 methods

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

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

The bottleneck: scalability

The scaling bottleneck

A determined quadratic interpolation model in $\mathbb{R}^n$ needs

$$\frac{(n+1)(n+2)}{2} \quad \text{sample points}$$

n	1	10	100	1000
sample points	3	66	5151	501501

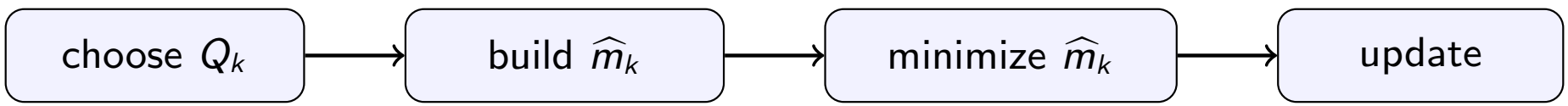
Full-space quadratic models become impractical for high-dimensional problems ($n \approx 1000$)

The main idea: model only in subspaces

Idea

At iteration k, replace a hard n-dimensional modeling problem with a lower-dimensional one on

$$x_k + Q_k \mathbb{R}^p, \qquad p \ll n$$



Note: A model built in the reduced coordinates can be lifted naturally back to the original space

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 $\mathbb{R}^n$:

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

Evaluate $f(x_k + s_k)$ and apply descent ratio test

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

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

Random subspace MBTR algorithm [Cartis, Roberts, 2023]

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 $\mathbb{R}^p$:

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

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

Fully linear and Q-fully linear models

Definition

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

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

Definition

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

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

Constructing Q-fully linear models

Let D consist of sample directions and write $D = QR$, where $Q^\top Q = I_p$

Theorem

If $f \in \mathcal{C}^{1+}$ and D is invertible, then the full space determined linear interpolation model is fully linear with $\kappa_f, \kappa_g = \mathcal{O}(\|(D/\max\|d_i\|)^{-1}\|)$

Theorem

If $f \in \mathcal{C}^{1+}$ and D has full column rank, then the subspace determined linear interpolation model is Q-fully linear with $\kappa_f, \kappa_g = \mathcal{O}(\|(R/\max\|r_i\|)^{-1}\|)$

Note: The $\|(D/\max\|d_i\|)^{-1}\|$ or $\|(R/\max\|r_i\|)^{-1}\|$ correspond to the geometry of the sample set

Quadratic models? $\Rightarrow$ QARSTA: Quadratic Approximation Random Subspace Trust-region Algorithm [Chen, Hare, Wiebe, 2024]

α-well-aligned matrices

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,*

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

Outline

Problem with constraints

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

General framework of 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 $\mathbb{R}^p$:

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

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

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*

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

  (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]

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 $n = 1000$:

- In some problems, COBYLA uses less fevals to reach f^*
- COBYLA requires more than 1000 times more time than CLARSTA to reach f^* (1min vs. 17h)

Results when $n = 10000$:

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

Outline

Summary

In summary,

- High-dimensional DFO problems can be addressed efficiently by random subspace methods
- QARSTA is a random subspace MBTR method for unconstrained problems, using low-dimensional quadratic models
- CLARSTA extends the unconstrained framework to convex constraints

Possible future directions:

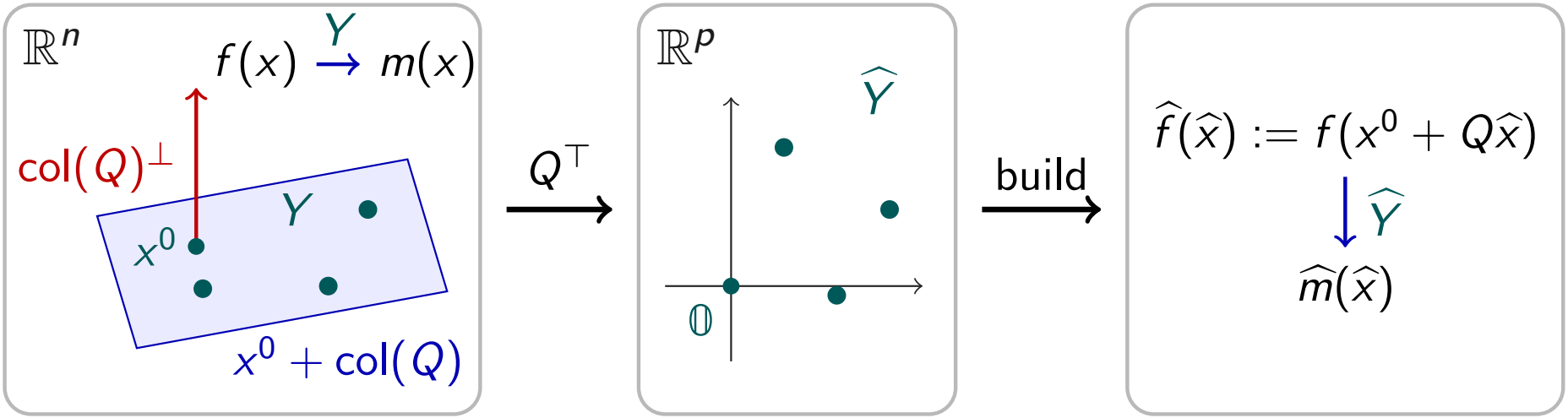
- Quadratic models in CLARSTA
- Better complexity bounds for CLARSTA
- Nonconvex/blackbox constraints

Thank you

- Yiwen Chen, Warren Hare, and Amy 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
- 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

Two equivalent views of subspace models



Lift and equivalence (Chen 2026)

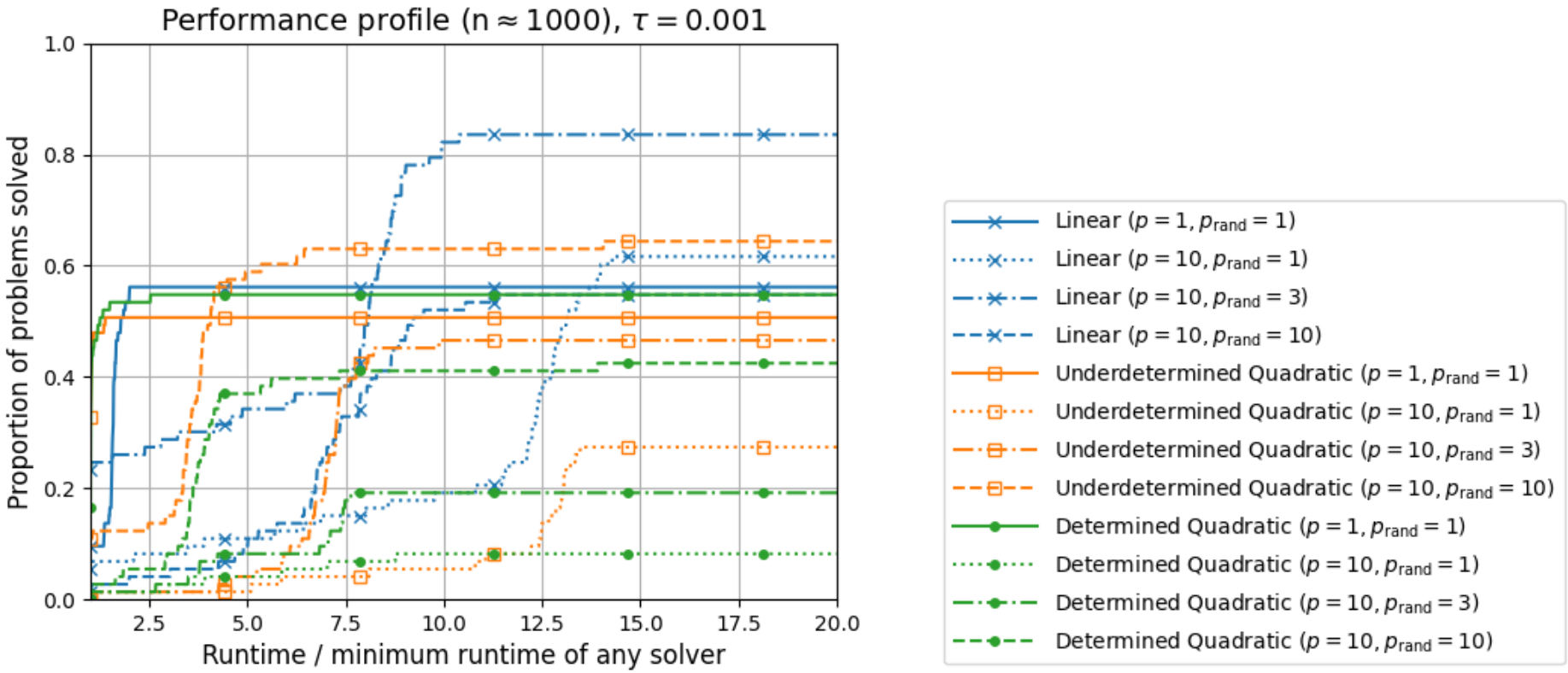
For many widely-used linear/quadratic models, we have

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

and

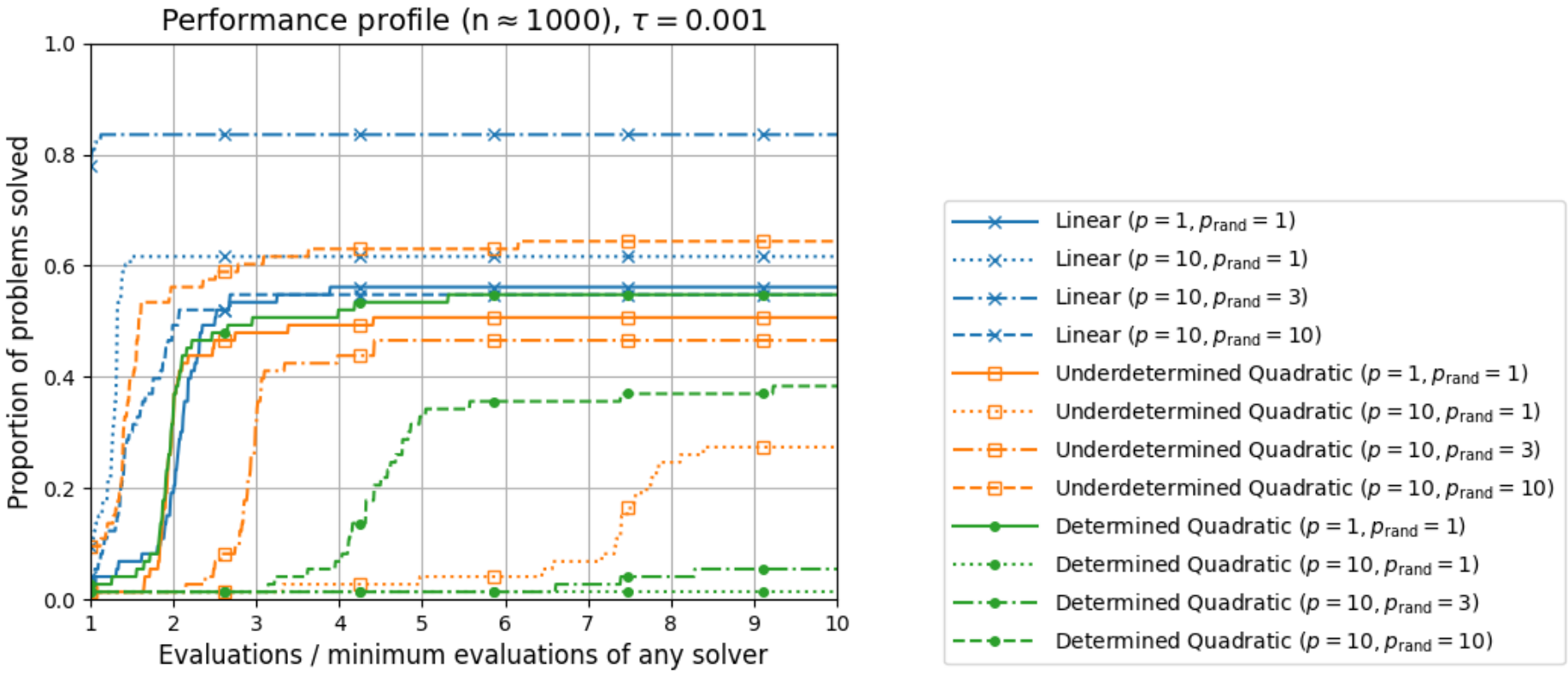
$$m(x) = \widehat{m}(\widehat{x}), \quad \text{for all } x \in \left(x^0 + Q\widehat{x}\right) + \mathrm{col}(Q)^\perp$$

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 f-evals



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