

A stochastic Levenberg-Marquardt method using random models

Clément Royer (*Université Paris Dauphine-PSL*)

Joint work with Elhoucine Bergou, Youssef Diouane, Vyacheslav Kungurtsev

EUROPT - June 9, 2021

Dauphine
UNIVERSITÉ PARIS

| PSL 

PR[AI]RIE
PaRis Artificial Intelligence Research InstitutE

Strongly constrained 4DVAR formulation

$$\begin{aligned} \min_{\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_T \in \mathbb{R}^n} & \quad \underbrace{\frac{1}{2} \|\mathbf{z}_0 - \mathbf{z}_b\|_{\mathbf{B}^{-1}}^2}_{\text{Background state}} + \underbrace{\frac{1}{2} \sum_{i=0}^T \|\mathbf{y}_i - \mathcal{H}(\mathbf{z}_i)\|_{\mathbf{R}_i^{-1}}^2}_{\text{Measurement fit}} \\ \text{s.t.} & \quad \underbrace{\mathbf{z}_i = \mathcal{M}_i(\mathbf{z}_{i-1}), \quad i = 1, \dots, T.}_{\text{Governing equations}} \end{aligned}$$

Strongly constrained 4DVAR formulation

$$\begin{aligned} \min_{\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_T \in \mathbb{R}^n} & \quad \underbrace{\frac{1}{2} \|\mathbf{z}_0 - \mathbf{z}_b\|_{\mathbf{B}^{-1}}^2}_{\text{Background state}} + \underbrace{\frac{1}{2} \sum_{i=0}^T \|\mathbf{y}_i - \mathcal{H}(\mathbf{z}_i)\|_{\mathbf{R}_i^{-1}}^2}_{\text{Measurement fit}} \\ \text{s.t.} & \quad \underbrace{\mathbf{z}_i = \mathcal{M}_i(\mathbf{z}_{i-1}), \quad i = 1, \dots, T.}_{\text{Governing equations}} \end{aligned}$$

Reformulation ($\mathbf{x} \leftrightarrow \mathbf{z}_0$)

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) := \frac{1}{2} \left(\|\mathbf{x} - \mathbf{z}_b\|_{\mathbf{B}^{-1}}^2 + \|\mathbf{y} - \mathcal{H}(\mathbf{x})\|_{\mathbf{R}^{-1}}^2 \right).$$

- Derivatives expensive to compute;
- \mathbf{B} unknown \Rightarrow Estimated via random samples.

A framework for nonlinear least squares

- Derivative-free context;
- Complexity analysis.

Probabilistic versions

- Randomly accurate models;
- Random function estimates.

And more recent work

- Exploit least-squares structure;
- More general complexity analysis.

- 1 Derivative-free least squares
- 2 Introducing randomness
- 3 New complexity results

- 1 Derivative-free least squares
- 2 Introducing randomness
- 3 New complexity results

Derivative-free nonlinear least-squares

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) := \frac{1}{2} \|\mathbf{r}(\mathbf{x})\|^2$$

- $\mathbf{r} : \mathbb{R}^n \mapsto \mathbb{R}^m$, $\mathbf{r} \in \mathcal{C}^{1,1}$;
- $\mathbf{J}(\mathbf{x}) := [\nabla r_i(\mathbf{x})^\top] \in \mathbb{R}^{m \times n}$ **not available**;
- Values of \mathbf{r} only accessed through **noisy estimates**.

Derivative-free nonlinear least-squares

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) := \frac{1}{2} \|\mathbf{r}(\mathbf{x})\|^2$$

- $\mathbf{r} : \mathbb{R}^n \mapsto \mathbb{R}^m$, $\mathbf{r} \in \mathcal{C}^{1,1}$;
- $\mathbf{J}(\mathbf{x}) := [\nabla r_i(\mathbf{x})^\top] \in \mathbb{R}^{m \times n}$ **not available**;
- Values of \mathbf{r} only accessed through **noisy estimates**.

Classical Levenberg-Marquardt approach

- Gauss-Newton model $f(\mathbf{x} + \mathbf{s}) \approx \frac{1}{2} \|\mathbf{r}(\mathbf{x}) + \mathbf{J}(\mathbf{x})\mathbf{s}\|^2 + \frac{\gamma}{2} \|\mathbf{s}\|^2$;
- **Regularization parameter** γ set adaptively.

Derivative-free Levenberg-Marquardt for $\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{r}(\mathbf{x})\|^2$

Inputs: $\mathbf{x}_0 \in \mathbb{R}^n, \mu_0 > 0, \eta > 0$.

Iteration j : Given (\mathbf{x}_j, μ_j) ,

- Compute $\mathbf{r}_{m_j} = \mathbf{r}(\mathbf{x}_j)$, $\mathbf{J}_{m_j} \approx \mathbf{J}(\mathbf{x}_j)$ and

$$\mathbf{s}_j \approx \operatorname{argmin}_{\mathbf{s}} m_j(\mathbf{s}) := \frac{1}{2} \|\mathbf{r}_{m_j} + \mathbf{J}_{m_j} \mathbf{s}\|^2 + \frac{\mu_j \|\mathbf{J}_{m_j}^\top \mathbf{r}_{m_j}\|}{2} \|\mathbf{s}\|^2.$$

- Compute $\mathbf{r}(\mathbf{x}_j)$ and $\mathbf{r}(\mathbf{x}_j + \mathbf{s}_j)$.

- If $\frac{\frac{1}{2} \|\mathbf{r}(\mathbf{x}_j)\|^2 - \frac{1}{2} \|\mathbf{r}(\mathbf{x}_j + \mathbf{s}_j)\|^2}{m_j(0) - m_j(\mathbf{s}_j)} \geq \eta$, set $\mathbf{x}_{j+1} = \mathbf{x}_j + \mathbf{s}_j$ and $\mu_{j+1} = 0.5\mu_j$;

- Otherwise, set $\mathbf{x}_{j+1} = \mathbf{x}_j$ and $\mu_{j+1} = 2\mu_j$.

- Regularization formula for **complexity results**.

Derivative-free Levenberg-Marquardt for $\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{r}(\mathbf{x})\|^2$

Inputs: $\mathbf{x}_0 \in \mathbb{R}^n, \mu_0 > 0, \eta > 0$.

Iteration j : Given (\mathbf{x}_j, μ_j) ,

- Compute $\mathbf{r}_{m_j} \approx \mathbf{r}(\mathbf{x}_j)$, $\mathbf{J}_{m_j} \approx \mathbf{J}(\mathbf{x}_j)$ and

$$\mathbf{s}_j \approx \operatorname{argmin}_{\mathbf{s}} m_j(\mathbf{s}) := \frac{1}{2} \|\mathbf{r}_{m_j} + \mathbf{J}_{m_j} \mathbf{s}\|^2 + \frac{\mu_j \|\mathbf{J}_{m_j}^\top \mathbf{r}_{m_j}\|}{2} \|\mathbf{s}\|^2.$$

- Compute $\mathbf{r}_j^0 \approx \mathbf{r}(\mathbf{x}_j)$ and $\mathbf{r}_j^s \approx \mathbf{r}(\mathbf{x}_j + \mathbf{s}_j)$.

- If $\frac{\frac{1}{2} \|\mathbf{r}_j^0\|^2 - \frac{1}{2} \|\mathbf{r}_j^s\|^2}{m_j(0) - m_j(\mathbf{s})} \geq \eta$, set $\mathbf{x}_{j+1} = \mathbf{x}_j + \mathbf{s}_j$ and $\mu_{j+1} = 0.5\mu_j$;

- Otherwise, set $\mathbf{x}_{j+1} = \mathbf{x}_j$ and $\mu_{j+1} = 2\mu_j$.

- Regularization formula for **complexity results**.

- Need **accuracy properties** for the models/estimates.

Recall: $m_j(\mathbf{s}) = \frac{1}{2} \|\mathbf{r}_{m_j} + \mathbf{J}_{m_j} \mathbf{s}\|^2 + \frac{\mu_j \|\mathbf{J}_{m_j}^\top \mathbf{r}_{m_j}\|}{2} \|\mathbf{s}\|^2$.

First-order accuracy

A model $m_j(\mathbf{s})$ is called κ_m -accurate if

$$\|\mathbf{J}(\mathbf{x}_j)^\top \mathbf{r}(\mathbf{x}_j) - \mathbf{J}_{m_j}^\top \mathbf{r}_{m_j}\| \leq \frac{\kappa_m}{\mu_j} \quad \text{and} \quad \left| \frac{1}{2} \|\mathbf{r}(\mathbf{x}_j)\|^2 - \frac{1}{2} \|\mathbf{r}_{m_j}\|^2 \right| \leq \frac{\kappa_m}{\mu_j^2}.$$

Model and estimates quality

Recall: $m_j(\mathbf{s}) = \frac{1}{2} \|\mathbf{r}_{m_j} + \mathbf{J}_{m_j} \mathbf{s}\|^2 + \frac{\mu_j \|\mathbf{J}_{m_j}^\top \mathbf{r}_{m_j}\|}{2} \|\mathbf{s}\|^2$.

First-order accuracy

A model $m_j(\mathbf{s})$ is called κ_m -accurate if

$$\|\mathbf{J}(\mathbf{x}_j)^\top \mathbf{r}(\mathbf{x}_j) - \mathbf{J}_{m_j}^\top \mathbf{r}_{m_j}\| \leq \frac{\kappa_m}{\mu_j} \quad \text{and} \quad \left| \frac{1}{2} \|\mathbf{r}(\mathbf{x}_j)\|^2 - \frac{1}{2} \|\mathbf{r}_{m_j}\|^2 \right| \leq \frac{\kappa_m}{\mu_j^2}.$$

Accurate function estimates

Estimates $(\mathbf{r}_j^0, \mathbf{r}_j^s)$ are κ_f -accurate if

$$\left| \frac{1}{2} \|\mathbf{r}_j^0\|^2 - \frac{1}{2} \|\mathbf{r}(\mathbf{x}_j)\|^2 \right| \leq \frac{\kappa_f}{\mu_j^2} \quad \text{and} \quad \left| \frac{1}{2} \|\mathbf{r}_j^s\|^2 - \frac{1}{2} \|\mathbf{r}(\mathbf{x}_j + \mathbf{s}_j)\|^2 \right| \leq \frac{\kappa_f}{\mu_j^2}.$$

Complexity analysis without derivatives

Goal: Bound the number of iterations to reach an ϵ_d -point \mathbf{x} such that

$$\|\nabla f(\mathbf{x})\| = \|\mathbf{J}(\mathbf{x})^\top \mathbf{r}(\mathbf{x})\| \leq \epsilon_d$$

as a function of ϵ_d .

Complexity analysis without derivatives

Goal: Bound the number of iterations to reach an ϵ_d -point \mathbf{x} such that

$$\|\nabla f(\mathbf{x})\| = \|\mathbf{J}(\mathbf{x})^\top \mathbf{r}(\mathbf{x})\| \leq \epsilon_d$$

as a function of ϵ_d .

Theorem (BDKR '18)

Suppose that

- Every model m_j is κ_m -accurate.
- Every estimate pair $(\mathbf{r}_j^0, \mathbf{r}_j^s)$ is κ_f -accurate.

Then, the method reaches an ϵ_d -point in at most

$$\mathcal{O}(\kappa^2 \epsilon_d^{-2})$$

iterations with $\kappa = \max\{1, \kappa_m, \kappa_f\}$.

- Derivative-based case: $\mathcal{O}(\epsilon_d^{-2})$.

- 1 Derivative-free least squares
- 2 Introducing randomness**
- 3 New complexity results

Recall : Our function estimates are noisy.

Inputs: $\mathbf{x}_0 \in \mathbb{R}^n, \mu_0 > 0$.

Iteration j : Given (\mathbf{x}_j, μ_j) ,

- Compute $\mathbf{r}_{m_j} \approx \mathbf{r}(\mathbf{x}_j)$, $\mathbf{J}_{m_j} \approx \mathbf{J}(\mathbf{x}_j)$ and
$$\mathbf{s}_j \approx \operatorname{argmin}_{\mathbf{s}} m_j(\mathbf{s}) = \frac{1}{2} \|\mathbf{r}_{m_j} + \mathbf{J}_{m_j} \mathbf{s}\|^2 + \frac{\mu_j \|\mathbf{J}_{m_j}^\top \mathbf{r}(\mathbf{x}_j)\|}{2} \|\mathbf{s}\|^2.$$
- Compute $\mathbf{r}_j^0 \approx \mathbf{r}(\mathbf{x}_j)$ and $\mathbf{r}_j^s \approx \mathbf{r}(\mathbf{x}_j + \mathbf{s}_j)$.
- If $\frac{\frac{1}{2} \|\mathbf{r}_j^0\|^2 - \frac{1}{2} \|\mathbf{r}_j^s\|^2}{m_j(0) - m_j(s)} \geq \eta$, set $\mathbf{x}_{j+1} = \mathbf{x}_j + \mathbf{s}_j$ and $\mu_{j+1} = 0.5\mu_j$;
- Otherwise, set $\mathbf{x}_{j+1} = \mathbf{x}_j$ and $\mu_{j+1} = 2\mu_j$.

- **Two sources of randomness** (models/estimates);
- Accounted for via **martingale-type properties**.

Accuracy property

For any realization, the model m_j is called κ_m -accurate if

$$\|\mathbf{J}(\mathbf{x}_j)^\top \mathbf{r}(\mathbf{x}_j) - \mathbf{J}_{m_j}^\top \mathbf{r}_{m_j}\| \leq \frac{\kappa_m}{\mu_j}, \quad \left| \frac{1}{2} \|\mathbf{r}(\mathbf{x}_j)\|^2 - \frac{1}{2} \|\mathbf{r}_{m_j}\|^2 \right| \leq \frac{\kappa_m}{\mu_j^2}$$

Accuracy property

For any realization, the model m_j is called κ_m -accurate if

$$\|\mathbf{J}(\mathbf{x}_j)^\top \mathbf{r}(\mathbf{x}_j) - \mathbf{J}_{m_j}^\top \mathbf{r}_{m_j}\| \leq \frac{\kappa_m}{\mu_j}, \quad \left| \frac{1}{2} \|\mathbf{r}(\mathbf{x}_j)\|^2 - \frac{1}{2} \|\mathbf{r}_{m_j}\|^2 \right| \leq \frac{\kappa_m}{\mu_j^2}$$

Probabilistic accuracy property

The **random** model sequence $\{m_j\}$ is called (p, κ_m) -accurate if

$$\forall j, \quad \mathbb{P}(m_j \text{ } \kappa_m\text{-accurate} \mid \mathcal{F}_{j-1}) \geq p.$$

- $\mathcal{F}_{j-1} = \sigma(m_0, \dots, m_{j-1}, \mathbf{r}_0^0, \mathbf{r}_0^s, \dots, \mathbf{r}_{j-1}^0, \mathbf{r}_{j-1}^s)$ represents the history of the algorithm up to iteration j .

Accurate function estimates

Estimates $(\mathbf{r}_j^0, \mathbf{r}_j^s)$ are κ_f -accurate if

$$\left| \frac{1}{2} \|\mathbf{r}_j^0\|^2 - \frac{1}{2} \|\mathbf{r}(\mathbf{x}_j)\|^2 \right| \leq \frac{\kappa_f}{\mu_j^2} \quad \text{and} \quad \left| \frac{1}{2} \|\mathbf{r}_j^s\|^2 - \frac{1}{2} \|\mathbf{r}(\mathbf{x}_j + \mathbf{s}_j)\|^2 \right| \leq \frac{\kappa_f}{\mu_j^2}.$$

Accurate function estimates

Estimates $(\mathbf{r}_j^0, \mathbf{r}_j^s)$ are κ_f -accurate if

$$\left| \frac{1}{2} \|\mathbf{r}_j^0\|^2 - \frac{1}{2} \|\mathbf{r}(\mathbf{x}_j)\|^2 \right| \leq \frac{\kappa_f}{\mu_j^2} \quad \text{and} \quad \left| \frac{1}{2} \|\mathbf{r}_j^s\|^2 - \frac{1}{2} \|\mathbf{r}(\mathbf{x}_j + \mathbf{s}_j)\|^2 \right| \leq \frac{\kappa_f}{\mu_j^2}.$$

Probabilistically accurate estimates

The **random** estimate sequence $\{(\mathbf{r}_j^0, \mathbf{r}_j^s)\}$ is (q, κ_f) -accurate if

$$\forall j, \quad \mathbb{P}((\mathbf{r}_j^0, \mathbf{r}_j^s) \text{ } \kappa_f\text{-accurate} \mid \mathcal{F}_{j-1/2}) \geq q.$$

- $\mathcal{F}_{j-1/2} = \sigma(m_0, \dots, m_{j-1}, m_j, \mathbf{r}_0^0, \mathbf{r}_0^s, \dots, \mathbf{r}_{j-1}^0, \mathbf{r}_{j-1}^s)$ represents the iteration of the algorithm up to the computation of \mathbf{r}_j^0 and \mathbf{r}_j^s .

Goal: Bound the **stopping time** $T_\epsilon = \min\{j \mid \|\mathbf{J}(\mathbf{x}_j)^\top \mathbf{r}(\mathbf{x}_j)\| \leq \epsilon_d\}$.

Goal: Bound the **stopping time** $T_\epsilon = \min\{j \mid \|\mathbf{J}(\mathbf{x}_j)^\top \mathbf{r}(\mathbf{x}_j)\| \leq \epsilon_d\}$.

Theorem (BDKR '18)

If $\{m_j\}$ is (p, κ_m) -accurate and $\{(\mathbf{r}_j^0, \mathbf{r}_j^s)\}$ is (q, κ_f) -accurate, then

$$\mathbb{E}[T_\epsilon] \leq \mathcal{O}\left(\frac{pq}{pq - 1/2} \kappa^2 \epsilon_d^{-2}\right),$$

$$\kappa = \max\{1, \kappa_m, \kappa_f\}.$$

Theorem (BDKR '20)

If for every iteration j , B is approximated using an ensemble of

- $n_m = \mathcal{O} \left(\max \left\{ \frac{\|\mathbf{x}_j - \mathbf{z}_b\|_{\mathbf{B}^{-1}}^2}{\kappa_m}, \frac{\|\mathbf{B}^{-1}(\mathbf{x}_j - \mathbf{z}_b)\|}{\kappa_m} \right\} n \right)$ samples to build m_j ,
- $n_f = \mathcal{O} \left(\frac{\|\mathbf{x}_j - \mathbf{z}_b\|_{\mathbf{B}^{-1}}^2}{\kappa_f} n \right)$ samples to compute $(\mathbf{r}_j^0, \mathbf{r}_j^s)$,

then there exists (p, q) such that $\{m_j\}$ is (p, κ_m) -accurate, and $\{(\mathbf{r}_j^0, \mathbf{r}_j^s)\}$ is (q, κ_f) -accurate.

- p, q depend on variance terms, likely unknown;
- $p, q \rightarrow 1$ as $n_m, n_f \rightarrow \infty$.

- 1 Derivative-free least squares
- 2 Introducing randomness
- 3 New complexity results**

About the complexity results

Our problem: $\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{r}(\mathbf{x})\|^2$

- Used $\|\mathbf{J}(\mathbf{x})^T \mathbf{r}(\mathbf{x})\|$ as a complexity metric;
- Oblivious to the least-square structure;
- May want to stop when residuals are small.

About the complexity results

Our problem: $\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{r}(\mathbf{x})\|^2$

- Used $\|\mathbf{J}(\mathbf{x})^T \mathbf{r}(\mathbf{x})\|$ as a complexity metric;
- Oblivious to the least-square structure;
- May want to stop when residuals are small.

Scaled gradient (Cartis, Gould, Toint '13; Gould, Rees, Scott '19)

$$\mathbf{g}(\mathbf{x}) := \begin{cases} \frac{\mathbf{J}(\mathbf{x})^T \mathbf{r}(\mathbf{x})}{\|\mathbf{r}(\mathbf{x})\|} & \text{if } \|\mathbf{r}(\mathbf{x})\| > 0 \\ 0 & \text{otherwise.} \end{cases}$$

- Stopping criterion for complexity:

$$\|\mathbf{r}(\mathbf{x})\| \leq \epsilon_p \quad \text{or} \quad \|\mathbf{g}(\mathbf{x})\| \leq \epsilon_d.$$

About the complexity results

Our problem: $\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{r}(\mathbf{x})\|^2$

- Used $\|\mathbf{J}(\mathbf{x})^T \mathbf{r}(\mathbf{x})\|$ as a complexity metric;
- Oblivious to the least-square structure;
- May want to stop when residuals are small.

Scaled gradient (Cartis, Gould, Toint '13; Gould, Rees, Scott '19)

$$\mathbf{g}(\mathbf{x}) := \begin{cases} \frac{\mathbf{J}(\mathbf{x})^T \mathbf{r}(\mathbf{x})}{\|\mathbf{r}(\mathbf{x})\|} & \text{if } \|\mathbf{r}(\mathbf{x})\| > 0 \\ 0 & \text{otherwise.} \end{cases}$$

- Stopping criterion for complexity:

$$\|\mathbf{r}(\mathbf{x})\| \leq \epsilon_p \quad \text{or} \quad \|\mathbf{g}(\mathbf{x})\| \leq \epsilon_d.$$

- Complexity of LM: for any $i \in \mathbb{N} \cup \{-1\}$, $\mathcal{O}(\epsilon_d^{-2} \epsilon_p^{-1/2^i})$.

New scaled gradient

Given $i \in \mathbb{N} \cup \{-1\}$,

$$\mathbf{g}^i(\mathbf{x}) := \begin{cases} \frac{\mathbf{J}(\mathbf{x})^T \mathbf{r}(\mathbf{x})}{\|\mathbf{r}(\mathbf{x})\|^{2-2^{-i}}} & \text{if } \|\mathbf{r}(\mathbf{x})\| \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

- Stopping criterion for complexity:

$$\|\mathbf{r}(\mathbf{x})\| \leq \epsilon_p \quad \text{or} \quad \|\mathbf{g}^i(\mathbf{x})\| \leq \epsilon_d.$$

New scaled gradient

Given $i \in \mathbb{N} \cup \{-1\}$,

$$\mathbf{g}^i(\mathbf{x}) := \begin{cases} \frac{\mathbf{J}(\mathbf{x})^T \mathbf{r}(\mathbf{x})}{\|\mathbf{r}(\mathbf{x})\|^{2-2^{-i}}} & \text{if } \|\mathbf{r}(\mathbf{x})\| \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

- Stopping criterion for complexity:

$$\|\mathbf{r}(\mathbf{x})\| \leq \epsilon_p \quad \text{or} \quad \|\mathbf{g}^i(\mathbf{x})\| \leq \epsilon_d.$$

- $i = -1$: Classical gradient;
- $i = 0$: CGT scaled gradient;
- $i \rightarrow \infty$: Resembles gradient dominance.

Complexity table

Goal: Find an (ϵ_p, ϵ_d) -point \mathbf{x}_k such that

$$\|\mathbf{r}(\mathbf{x}_k)\| \leq \epsilon_p \quad \text{or} \quad \|\mathbf{g}^i(\mathbf{x}_k)\| \leq \epsilon_d.$$

Complexity results (BDKR '20)

i	Arbitrary	$i = -1$	$i = 0$	$i \rightarrow \infty$
$\ \mathbf{g}^i(\mathbf{x})\ $	$\frac{\ \mathbf{J}(\mathbf{x})^T \mathbf{r}(\mathbf{x})\ }{\ \mathbf{r}(\mathbf{x})\ ^{2-2^{-i}}}$	$\ \mathbf{J}(\mathbf{x})^T \mathbf{r}(\mathbf{x})\ $	$\frac{\ \mathbf{J}(\mathbf{x})^T \mathbf{r}(\mathbf{x})\ }{\ \mathbf{r}(\mathbf{x})\ }$	$\frac{\ \mathbf{J}(\mathbf{x})^T \mathbf{r}(\mathbf{x})\ }{\ \mathbf{r}(\mathbf{x})\ ^2}$
Order	$\epsilon_d^{-2} \epsilon_p^{-(4-2^{1-i})}$	ϵ_d^{-2}	$\epsilon_d^{-2} \epsilon_p^{-2}$	$\epsilon_d^{-2} \epsilon_p^{-4}$.

- Probabilistic versions (expected stopping time);
- For the connoisseur: other analyzes get better bounds in terms of ϵ_p/ϵ_d but have exponential dependencies in 2^i .

Our contributions

- Redefined probabilistic property for Levenberg-Marquardt schemes (analogy with trust region).
- Complexity analysis for **stochastic function estimates/stochastic models**.
- A family of **complexity metrics** and results.

Derivative-free Levenberg-Marquardt

A stochastic Levenberg-Marquardt method using random models with complexity results.

E. Bergou, Y. Diouane, V. Kungurtev and C. W. Royer.
<https://arXiv.org/abs/1807.02176v3> coming soon!

A nonmonotone matrix-free algorithm for nonlinear equality-constrained least-squares problems.

E. Bergou, Y. Diouane, V. Kungurtev and C. W. Royer.
SIAM Journal on Scientific Computing.

- Addressed (among others) the original 4Dvar formulation:

$$\min_{\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_T \in \mathbb{R}^n} \quad \frac{1}{2} \|\mathbf{z}_0 - \mathbf{z}_b\|_{\mathbf{B}^{-1}}^2 + \frac{1}{2} \sum_{i=0}^T \|\mathbf{y}_i - \mathcal{H}(\mathbf{z}_i)\|_{\mathbf{R}_i^{-1}}^2$$

$$\text{s.t.} \quad \mathbf{z}_i = \mathcal{M}_i(\mathbf{z}_{i-1}), \quad i = 1, \dots, T.$$

- Derivative-based approach but **inexact**.

Going forward

- Combine random models with constraints;
- Practical probabilistic approaches.

PR[AI]RIE

PaRis Artificial Intelligence Research InstitutE

Opening for a postdoc

- Context: My springboard chair position on optimization;
- Two years, start date between Sep. 2021-Jan. 2022;
- Research topics include derivative-free optimization for automated ML!

PR[AI]RIE

PaRis Artificial Intelligence Research InstitutE

Opening for a postdoc

- Context: My springboard chair position on optimization;
- Two years, start date between Sep. 2021-Jan. 2022;
- Research topics include derivative-free optimization for automated ML!

Wanna know more?

Email me! clement.royer@dauphine.psl.eu

PR[AI]RIE

PaRis Artificial Intelligence Research InstitutE

Opening for a postdoc

- Context: My springboard chair position on optimization;
- Two years, start date between Sep. 2021-Jan. 2022;
- Research topics include derivative-free optimization for automated ML!

Wanna know more?

Email me! clement.royer@dauphine.psl.eu

Thank you all!