Recent Results in Infinite-Dimensional Optimization

Gökhan Göksu[a]
(Joint Work with Alessio Moreschini[b] and Thomas Parisini[b,c,d])

[a]Department of Mathematical Engineering,
Yıldız Technical University, Türkiye
[b]Department Electrical and Electronic Engineering,
Imperial College London, London, UK
[c]KIOS Research and Innovation Center of Excellence,
University of Cyprus, Nicosia, Cyprus
[d]Dipartimento Ingegneria e Architettura,
Università di Trieste, Trieste, Italy

MTM4502 - Optimization Techniques / Research Seminar

Introduction

Discrete Gradient in Finite-Dimensional Context

A **discrete gradient** is a continuous operator $\bar{\nabla} S : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ such that for all pairs $x_1, x_2 \in \mathbb{R}^n$

$$\langle x_1 - x_2, \bar{\nabla} S(x_1, x_2) \rangle = S(x_1) - S(x_2), \quad \text{(Mean Value Property)}$$
$$\lim_{x_2 \to x_1} \bar{\nabla} S(x_1, x_2) = \nabla S(x_1), \quad \text{(Continuity Property)}$$

- ∇ is the Euclidean gradient (vector differential operator),
- $\langle \cdot, \cdot \rangle$ is the standard Euclidean inner product.
- Introduced by Gonzalez (1996).
- Address loss of energy conservation in numerical solutions
- Multiple discrete gradient operators proposed for higher-dimensional spaces: Gonzalez (1996), Harten et. al. (1983), Itoh and Abe (1988), Moreschini et al. (2024).

Introduction

Significance in Numerical Optimization and Challenges in Infinite-Dimensional Spaces

Significance:

- Observed discrepancy between continuous and discrete dynamics
- Discrete gradient maintains properties like energy conservation and Lyapunov functions
- Stability in optimization methods
- Finite-dimensional discrete gradient method for convex optimization

Challenges:

- Optimization Over Infinite-Dimensional Spaces
- Lack of formal definition of discrete gradient for infinite-dimensional spaces
- Sensitivity of existing gradient-based methods to time step

Preliminaries

Notation and Definitions: Fréchet Derivative

- $\mathcal{C}([a,b],\mathbb{R}^n)$: set of continuous functions
- For any $\phi_1, \phi_2 \in \mathcal{C}([a,b],\mathbb{R}^n)$, the inner product is defined by

$$\langle \phi_1, \phi_2 \rangle := \int_a^b \phi_2^\top(s) \cdot \phi_1(s) ds.$$

Definition (Fréchet Derivative)

Any functional $S : \mathcal{C}([a,b],\mathbb{R}^n) \to \mathbb{R}$ is said to be Fréchet differentiable at $\phi_0 \in \mathcal{C}([a,b],\mathbb{R}^n)$ if there exists a bounded linear operator $\nabla_F S(\phi_0) : \mathcal{C}([a,b],\mathbb{R}^n) \to \mathbb{R}$ such that

$$\lim_{\|\phi\| \to 0^+} \frac{|S(\phi_0 + \phi) - S(\phi_0) - \nabla_F S(\phi_0)(\phi)|}{\|\phi\|} = 0.$$

The operator $\nabla_F S(\phi_0)$ is called the **Fréchet derivative evaluated at** $\phi_0 \in \mathcal{C}([a,b],\mathbb{R}^n)$, or simply Fréchet derivative when it is clear from the context.

Preliminaries

Definitions: Second Order Fréchet Derivative

Definition (Second Order Fréchet Derivative)

Any functional $S : \mathcal{C}([a,b],\mathbb{R}^n) \to \mathbb{R}$ is said to be twice Fréchet differentiable at $\phi_0 \in \mathcal{C}([a,b],\mathbb{R}^n)$ if there exists a bounded linear operator $\nabla_F^2 S(\phi_0) : \mathcal{C}([a,b],\mathbb{R}^n) \times \mathcal{C}([a,b],\mathbb{R}^n) \to \mathbb{R}$ uniformly for $\phi_1 \in \mathcal{C}([a,b],\mathbb{R}^n)$ on bounded sets of $\mathcal{C}([a,b],\mathbb{R}^n)$ such that

$$\lim_{\|\phi_2\|\to 0^+} \frac{1}{\|\phi_2\|}\Big(|\nabla_F S(\phi_0+\phi_2)(\phi_1) - \nabla_F S(\phi_0)(\phi_1) - \nabla_F^2 S(\phi_0)(\phi_1,\phi_2)|\Big) = 0.$$

The operator $\nabla_F^2 S(\phi_0)$ is called the **second order Fréchet derivative evaluated at** $\phi_0 \in \mathcal{C}([a,b],\mathbb{R}^n)$, or simply second order Fréchet derivative when it is clear from the context.

Preliminaries

Illustrative Example: Fréchet Derivatives

Consider the following functional, defined for all $\phi \in \mathcal{C}([a,b],\mathbb{R}^n)$,

$$S(\phi) = \|\phi\|^2 = \int_a^b |\phi(s)|^2 ds. \tag{4}$$

The first and the second order Fréchet derivatives of (4) are given for all $\varphi \in \mathcal{C}([a,b],\mathbb{R}^n)$ and for all $(\varphi_1,\varphi_2) \in \mathcal{C}([a,b],\mathbb{R}^n) \times \mathcal{C}([a,b],\mathbb{R}^n)$ by

$$\nabla_F S(\phi)(\varphi) = \int_a^b 2\varphi^\top(s)\phi(s)ds, \tag{5}$$

$$\nabla_F^2 S(\phi)(\varphi_1,\varphi_2) = \int_a^b 2\varphi_1^\top(s)\varphi_2(s)ds. \tag{6}$$

Lemma (Riesz Representation Theorem)

Let $\nabla_F S(\phi_0)$ be a bounded linear mapping defined as in Definition 1. Then, there exists a $\nabla_R S(\phi_0) \in \mathcal{C}([a,b],\mathbb{R}^n)$ such that the following holds, for every $\varphi \in \mathcal{C}([a,b],\mathbb{R}^n)$,

$$\nabla_F S(\phi_0)(\varphi) = \langle \nabla_R S(\phi_0), \varphi \rangle.$$

Lemma (Riesz “Hessian” Representation Theorem)

Let $\nabla_F^2 S(\phi_0)$ be a bounded linear operator defined as in Definition 2. Then, there exists a bounded linear operator $\nabla_R^2 S(\phi_0) : \mathcal{C}([a,b],\mathbb{R}^n) \to \mathcal{C}([a,b],\mathbb{R}^n)$ such that for every $\varphi_1, \varphi_2 \in \mathcal{C}([a,b],\mathbb{R}^n)$ we have

$$\nabla_F^2 S(\phi_0)(\varphi_1,\varphi_2) = \langle \nabla_R^2 S(\phi_0)(\varphi_2), \varphi_1 \rangle.$$

Preliminaries

Illustrative Example: Riesz “Gradient" and “Hessian" Representation Theorems

The Fréchet derivatives of the functional

$$S(\phi) = \|\phi\|^2 = \int_a^b |\phi(s)|^2 ds$$

were

$$\nabla_F S(\phi)(\varphi) = \int_a^b 2\varphi^\top(s)\phi(s)ds, \tag{5}$$

$$\nabla_F^2 S(\phi)(\varphi_1, \varphi_2) = \int_a^b 2\varphi_1^\top(s)\varphi_2(s)ds. \tag{6}$$

Here, invoking Riesz “Gradient" and “Hessian" Representation Theorems for (5) and (6), we obtain the representations

$$\nabla_R S(\phi) = 2\phi, \ \forall \phi \in \mathcal{C}([a, b], \mathbb{R}^n),$$

$$\nabla_R^2 S(\phi)(\varphi) = 2\varphi, \ \forall \phi, \varphi \in \mathcal{C}([a, b], \mathbb{R}^n).$$

Fréchet Discrete Operators

Fréchet Discrete Gradient and Hessian

The infinite-dimensional counterpart of the discrete gradient is understood as a discrete vector representation of the Fréchet derivative $\nabla_F S(\phi)$ in $\mathcal{C}([a,b],\mathbb{R}^n)$.

Definition (Fréchet Discrete Gradient)

Given a Fréchet differentiable functional $S : \mathcal{C}([a,b],\mathbb{R}^n) \to \mathbb{R}$ and $\nabla_R S(\phi_0) \in \mathcal{C}([a,b],\mathbb{R}^n)$, a **Fréchet discrete gradient** is a bounded linear operator $\bar{\nabla} S$ such that for all $\phi, \varphi \in \mathcal{C}([a,b],\mathbb{R}^n)$

$$\langle \phi - \varphi, \bar{\nabla} S(\phi,\varphi)\rangle = S(\phi) - S(\varphi),$$
$$\lim_{\varphi \to \phi} \bar{\nabla} S(\phi,\varphi) = \nabla_R S(\phi).$$

Definition (Fréchet Discrete Hessian)

Given a twice Fréchet differentiable functional $S : \mathcal{C}([a,b],\mathbb{R}^n) \to \mathbb{R}$ and $\nabla_R^2 S(\phi_0)$, a **Fréchet discrete Hessian** is a bounded linear ope-rator $\bar{\nabla}^2 S$ such that for all $\phi, \varphi, \psi \in \mathcal{C}([a,b],\mathbb{R}^n)$

$$S(\psi) - S(\varphi) = \langle \psi - \varphi, \bar{\nabla} S(\phi,\varphi) + \bar{\nabla}^2 S(\phi,\varphi,\psi)(\psi - \varphi)\rangle,$$
$$\lim_{\psi \to \varphi} \bar{\nabla}^2 S(\phi,\varphi,\psi) = \nabla_R^2 S(\phi)(\varphi).$$

Fréchet Discrete Operators

Main Result: Fréchet Discrete Gradient 1

Proposition

Let S be a Fréchet differentiable functional. Suppose there exist functionals V and W such that

$$
\begin{aligned}
i) &\quad \int_a^b V(\phi(s), \varphi(s))ds = S(\phi) - S(\varphi), \\
ii) &\quad \langle \phi - \varphi, W(\phi, \varphi) \rangle = 0, \\
iii) &\quad \lim_{\varphi \to \phi} \frac{V(\phi, \varphi)}{\|\phi - \varphi\|^2}(\phi - \varphi) = \nabla_R S(\phi) - W(\phi, \phi).
\end{aligned}
$$

Then, $\bar{\nabla} S(\phi, \varphi) = \frac{V(\phi,\varphi)}{\|\phi-\varphi\|^2}(\phi - \varphi) + W(\phi, \varphi)$ *is a Fréchet discrete gradient.*

Fréchet Discrete Operators

Main Result: Fréchet Discrete Hessian 1

Proposition

Let S be a twice Fréchet differentiable functional. Suppose there exist skew-symmetric operators K_1 and K_2 such that

$$\lim_{\psi \to \varphi} \Sigma(\phi, \varphi, \psi) = \nabla_R^2 S(\phi)(\varphi) - K_1(\varphi, \varphi),$$

$$\Sigma(\phi, \varphi, \psi) := \left(\bar{\nabla} S(\psi, \varphi) - \bar{\nabla} S(\phi, \varphi)\right) \frac{(\psi - \varphi)^\top}{\|\psi - \varphi\|^2} + \left(\bar{\nabla} S(\psi, \varphi) - \bar{\nabla} S(\phi, \varphi)\right) (\psi - \varphi)^\top K_2(\phi, \varphi, \psi).$$

Then, $\bar{\nabla}^2 S(\phi, \varphi, \psi) = K_1(\varphi, \psi) + \Sigma(\phi, \varphi, \psi)$ is a Fréchet discrete Hessian.

Fréchet Discrete Operators

Main Result: Fréchet Discrete Gradient 2 (For the Functionals in Integral Form)

Proposition

Let S be a Fréchet differentiable functional defined as

$$S(\phi) = \int_a^b \sigma(\phi(s))\, ds,$$

where $\sigma : \mathbb{R}^n \to \mathbb{R}$*. The bounded linear operator*

$$\bar{\nabla} S := (\nabla_{d_1} S, \cdots, \nabla_{d_n} S)^\top,$$

where $\nabla_{d_i} S(\phi, \psi)$ *is defined as*

$$\nabla_{d_1} S(\phi, \psi) = \frac{\sigma(\phi_1, \psi_2, \cdots, \psi_n) - \sigma(\psi_1, \cdots, \psi_n)}{\phi_1 - \psi_1},$$

$$\nabla_{d_n} S(\phi, \psi) = \frac{\sigma(\phi_1, \cdots, \phi_n) - \sigma(\phi_1, \cdots, \phi_{n-1}, \psi_n)}{\phi_n - \psi_n},$$

$$\nabla_{d_i} S(\phi, \psi) = \frac{\sigma(\phi_1, \cdots, \phi_i, \psi_{i+1}, \cdots, \psi_n) - \sigma(\phi_1, \cdots, \phi_{i-1}, \psi_i, \cdots, \psi_n)}{\phi_i - \psi_i},$$

is a Fréchet discrete gradient.

Fréchet Discrete Operators

Main Result: Fréchet Discrete Gradient 3

Proposition

Let S be a Fréchet differentiable functional. Assume that for any $\phi, \varphi \in \mathcal{C}([a,b], \mathbb{R}^n)$ and all $\xi \in (0,1)$, the convex combination $\xi\phi + (1-\xi)\varphi$ is contained in $\mathcal{C}([a,b], \mathbb{R}^n)$. Then, the bounded linear operator

$$\bar{\nabla} S(\phi, \varphi) = \begin{bmatrix} \int_0^1 [\nabla_R S(\xi\phi + (1-\xi)\varphi)]_1 \, d\xi \\ \vdots \\ \int_0^1 [\nabla_R S(\xi\phi + (1-\xi)\varphi)]_n \, d\xi \end{bmatrix},$$

is a Fréchet discrete gradient.

Fréchet Discrete Operators

Illustrative Example: Fréchet Discrete Gradient

Consider the functional

$$S(\phi) = \int_a^b (\phi_1(s) - f_1(s))^4 + (\phi_2(s) - f_2(s))^2 \, ds,$$

where $f_1, f_2 \in \mathcal{C}([a, b], \mathbb{R})$. The Fréchet derivative is

$$\nabla_R S(\phi) = \begin{bmatrix} 4(\phi_1 - f_1)^3 \\ 2(\phi_2 - f_2) \end{bmatrix}.$$

The Fréchet discrete gradient is given by

$$\bar{\nabla} S(\phi, \varphi) = \begin{bmatrix} \left((\phi_1 - f_1)^2 + (\varphi_1 - f_1)^2\right)(\phi_1 + \varphi_1 - 2f_1) \\ (\phi_2 + \varphi_2 - 2f_2) \end{bmatrix}.$$

Consider the unconstrained optimization problem

$$\min_{\phi \in \mathcal{C}([a,b],\mathbb{R}^n)} S(\phi),$$

where $S : \mathcal{C}([a,b],\mathbb{R}^n) \to \mathbb{R}$ is the objective functional.

Given an objective functional S and an initial guess $\phi_0 \in \mathcal{C}([a,b],\mathbb{R}^n)$, the **Fréchet discrete gradient method** is

$$\phi_{k+1} = \phi_k - \tau_k \bar{\nabla} S(\phi_{k+1}, \phi_k),$$

where $\tau_k > 0$ is a k-varying time step.

Definition

Let $\mathcal{X}$ be a Banach space. A functional $S : \mathcal{X} \to \mathbb{R}$ is **lower semi-continuous** at $\underline{\phi} \in \mathcal{X}$ if

$$S(\underline{\phi}) \leq \liminf_{k\to\infty} S(\phi)$$

for all sequences $\{\phi_k\}_{k\in\mathbb{N}} \in \mathcal{X}$ such that $\phi_k \to \underline{\phi}$.

Definition

A functional $S : \mathcal{X} \to \mathbb{R}$ is **coercive** if $\lim_{\|\phi\|\to\infty} S(\phi) = \infty$.

Definition

Let $\mathcal{X}$ be a non empty convex subset of $\mathcal{C}([a,b],\mathbb{R}^n)$. A functional $S : \mathcal{C}([a,b],\mathbb{R}^n) \to \mathbb{R}$ is **convex** if for all $\alpha \in [0,1]$ and for all $\phi, \varphi \in \mathcal{X}$

$$S(\alpha\phi + (1-\alpha)\varphi) \leq \alpha S(\phi) + (1-\alpha)S(\varphi).$$

The functional S is **strictly convex** if the above inequality holds tight for all $\phi \neq \varphi$ and $\alpha \in (0,1)$.

Theorem

Let $\mathcal{X}$ be a non-empty, convex, strongly closed, and bounded subset of $\mathcal{C}([a,b],\mathbb{R}^n)$. Suppose S is convex, lower semi-continuous, and Fréchet differentiable on every bounded subset of $\mathcal{X}$. Then any sequence $\{\phi_k\}$ generated by the Fréchet discrete gradient method is a (locally) minimizing sequence for S

$$S(\phi_k) \to \inf_{\mathcal{X}} S(\phi).$$

Moreover, for all $\phi_{k+1} \neq \phi_k$

$$S(\alpha\phi_{k+1} + (1-\alpha)\phi_k) \leq S(\phi_k).$$

Corollary

Suppose S is strictly convex, lower semi-continuous, and Fréchet differentiable on every bounded subset of $\mathcal{X}$. Then any sequence $\{\phi_k\}$ generated by the Fréchet discrete gradient method converges to the singleton set $\inf_{\mathcal{X}} S(\phi)$. Moreover, for all $\phi_{k+1} \neq \phi_k$

$$S(\alpha\phi_{k+1} + (1-\alpha)\phi_k) < S(\phi_k).$$

Infinite Dimensional Optimization

Illustrative Example

Consider the least-squares problem for an integral operator

$$\min_{\phi \in \mathcal{C}([a,b],\mathbb{R}^2)} S(\phi) = \int_a^b \sum_{i=1}^{2} (\phi_i(s) - f_i(s))^2 \, ds,$$

where $f_1, f_2 \in \mathcal{C}([a,b],\mathbb{R})$.

The Fréchet derivative for all $\varphi_1, \varphi_2 \in \mathcal{C}([a,b],\mathbb{R})$ is

$$\nabla_F S(\phi)(\varphi) = \int_a^b \sum_{i=1}^{2} 2(\phi_i(s) - f_i(s))\varphi_i(s) \, ds.$$

The representation of the Fréchet derivative is

$$\nabla_R S(\phi) = 2(\phi - f),$$

where

$$f = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}, \quad \phi = \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}.$$

The Fréchet discrete gradient method for the integral operator yields

$$\phi_{k+1} = \phi_k - \frac{2\tau_k}{1+\tau_k}(\phi_k - f).$$

Consider $a = -1$, $b = 1$, and for all $x \in [-1, 1]$

$$f = \begin{bmatrix} \sin(x) \\ \cos(x) \end{bmatrix}, \quad \phi_0 = \begin{bmatrix} \cos(4x) \\ \frac{4}{3} + \log(\sqrt[3]{x + 1.2}) \end{bmatrix}.$$

Infinite Dimensional Optimization

Illustrative Example: Error Analysis

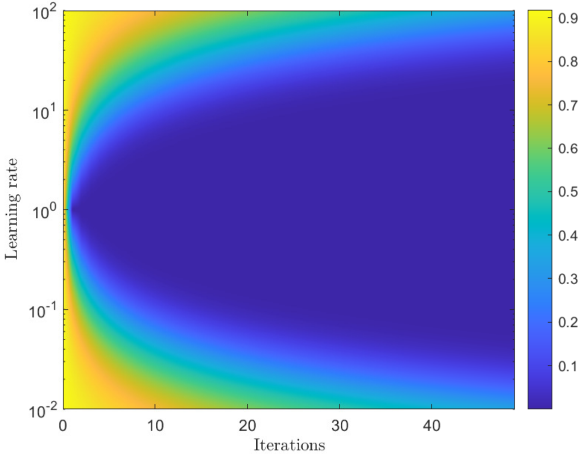

Figure: Contour of the error norm $\|\phi_k - f\|$ for all iterations $k \in [0, 50]$ and learning rates $\tau \in [10^{-2}, 10^2]$.

Infinite Dimensional Optimization

Illustrative Example: The Sequence $\{\phi_k\}$

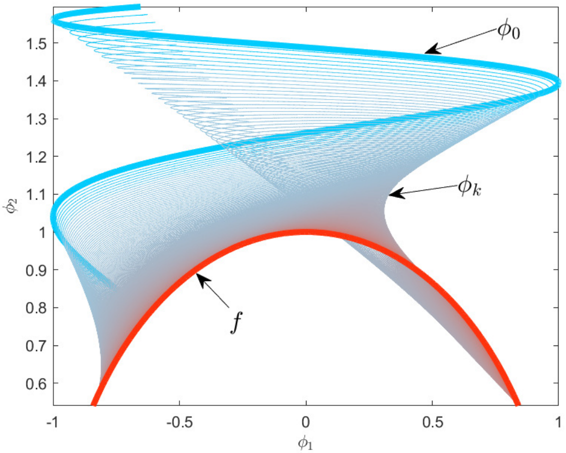

Figure: Shape of the sequence $\{\phi_k\}$ for all iterations $k \in [0, 200]$ and learning rate $\tau = 10^{-2}$.

Conclusion

- The Fréchet discrete gradient method extends discrete gradient methods to infinite-dimensional spaces, providing effective tools for optimization with convergence guarantees.
- This study presents a novel approach to leverage the representation of Fréchet derivatives on Banach spaces by introducing Fréchet discrete operators to infinite-dimensional spaces.
- The main contributions include:
  - **Fréchet Discrete Gradient:** An extension of the discrete gradient concept by González et al. (1996) for finite-dimensional spaces.
  - **Fréchet Discrete Hessian:** An enhancement of the second-order representation of the Fréchet derivative.
- We provide the first insight into discrete gradient methods for convex optimization on infinite-dimensional spaces.
- Under mild conditions, any sequence generated by the discrete gradient method built on Fréchet discrete gradient operators achieves convergence for all finite learning rates.

Future Work

Further exploration and application of Fréchet discrete operators can lead to significant advancements in:

- Optimization theory
- Functional learning
- Optimal control problems and team theory (Zoppoli et al., 2020, Ch. 9)
- Data-driven model reduction, see Moreschini et al. (2023a,c), Simard et al. (2023), Moreschini et al. (2024)

THANK YOU!
QUESTIONS?