A probabilistic result on impulsive noise reduction through Group Equivariant Non-Expansive Operators

Patrizio Frosini\textsuperscript{1,2,3,4}, Ivan Gridelli\textsuperscript{1} and Andrea Pascucci\textsuperscript{1}

\textsuperscript{1}Department of Mathematics, University of Bologna, Italy.\textsuperscript{2}ALMA-AI, University of Bologna, Italy.\textsuperscript{3}AM\textsuperscript{2}, University of Bologna, Italy.\textsuperscript{4}ARCES, University of Bologna, Italy.

Contributing authors: patrizio.frosini@unibo.it; ivangridelli@gmail.com; andrea.pascucci@unibo.it;

Abstract

In recent years, group equivariant non-expansive operators (GENEOs) have attracted attention in the fields of Topological Data Analysis and Machine Learning. In this paper we show how these operators can be of use also for the removal of impulsive noise and to increase the stability of TDA in the presence of noisy data. In particular, we prove that GENEOs can control the expected value of the perturbation of persistence diagrams caused by uniformly distributed impulsive noise, when data are represented by $L$-Lipschitz functions from $\mathbb{R}$ to $\mathbb{R}$.

Keywords: GENEO, impulsive noise, persistence diagram, persistent homology, machine learning

MSC Classification: MSC Primary: 55N31, 62R40; Secondary: 60-08, 65D18, 68T09, 68U05.

1 Introduction

In the last thirty years Topological Data Analysis (TDA) developed as a useful mathematical theory for analyzing data, benefiting from the reduction of dimensionality guaranteed by topology\cite{3, 5, 11}. One of the main tools in TDA
is the concept of persistence diagram, which is a collection of points in the real plane, describing the homological changes of the sublevel sets of suitable continuous functions. These changes give important information about the data of interest, focusing on some of their most relevant properties. Persistence diagrams can be used in the presence of noise, since a well-known stability theorem states that these topological descriptors change in a controlled way when we know that the functions expressing the filtrations we are interested in change in a controlled way with respect to the sup-norm \[8\]. Furthermore, \(L^p\)-stability of persistence diagrams with respect to the sup-norm have been proved in \[9\]. Unfortunately, in many applications the sup-norm of the noise is not guaranteed to be small, and hence these results cannot be directly applied. In particular, these results cannot be directly used when data are affected by impulsive noise.

Analogously, in the discrete setting of TDA the presence of outliers in cloud points can drastically affect the corresponding persistence diagrams. The problem of managing outliers has been studied by several authors by different techniques. In \[12\] an approach based on confidence sets has been introduced. A method inspired from the \(k\)-nearest neighbors regression and the local median filtering has been used in \[4\], while the concept of bagplot has been applied in \[1\]. In \[16\] an approach based on reproducing kernels has been proposed.

In our paper we start exploring a different probabilistic approach, in the topological setting. The main idea is that studying the properties of data is a process that should be primarily based on the analysis of the observers that are asked to examine the data, since we cannot ignore that different observers can differently judge the same data. This approach has been initially proposed in \[2\] and requires both the definition of the space of Group Equivariant Non-Expansive Operators (GENEOs) and the development of geometrical techniques to move around in this space \[10\]. In other words, in this model we should not wonder how we have to manage the data but rather how we have to manage the observers (i.e., GENEOs) analyzing the data. As an initial step in this direction, in this paper we show how GENEOs can be used to get stability of persistence diagrams of 1D-signals in the presence of impulsive noise.

GENEOs have been studied in \[13\] as a new tool in TDA, since they allow for an extension of the theory that is not invariant under the action of every homeomorphism of the considered domain. This is important in applications where the invariance group is not the group of all homeomorphisms, such as the ones concerning shape comparison. Interestingly, GENEOs are also deeply related to the foliation method used to define the matching distance in 2-dimensional persistent homology \[6, 7\] and can be seen as a theoretical bridge between TDA and Machine Learning \[2\]. Furthermore, these operators make available lower bounds for the natural pseudo-distance \(d_G(\varphi_1, \varphi_2) := \inf_{g \in G} \|\varphi_1 - \varphi_2 \circ g\|_\infty\), associated with a group \(G\) of self-homeomorphisms of the domain of the signals \(\varphi_1, \varphi_2\) \[13\].
In our paper, we prove that GENEOs can control the expected value of the perturbation of persistence diagrams caused by uniformly distributed impulsive noise, when data are represented by $L$-Lipschitz functions from $\mathbb{R}$ to $\mathbb{R}$. In order to do that, we choose a “mother” bump function $\psi$, i.e. a continuous function that is non-negative, upper bounded by 1 and compactly supported, and we assume that our noise is made up of finite bumps each obtained by translating, heightening and/or widening $\psi$. The function $\hat{\varphi} = \varphi + R$ represents the corrupted data, where the noise is given by the function $R = \sum_{i=1}^{k} a_i \psi(b_i(x - c_i))$ for some $a_i, b_i, c_i \in \mathbb{R}$, with $b_i > 0$, and some positive integer $k$.

In this situation, trying to use a convolution to approximate our starting data is not effective, because even if it does contract the bumps, it does not cut them, and hence does not improve the sup-norm distance from the original data. A classical approach for the removal of impulsive noise is the one of using a median filter[15]. Although this approach would be quite efficient in the discrete case, let us remark that in our setting we are considering continuous functions. The analogous of the median for the continuous case would be defined as the interval of those values $m$ such that

$$\int_{-\infty}^{m} f(x)dx = \int_{m}^{+\infty} f(x)dx$$

where $f$ is a density of probability. However, this operator is not stable: a small alteration of the starting function could lead to a significant change of the median.

The operators we consider in this paper are $F^\delta = \max(\varphi(x - \delta), \varphi(x + \delta))$ and $F_\varepsilon = \min(\varphi(x - \varepsilon), \varphi(x + \varepsilon))$. The main idea is that $F^\delta$ cuts the noise “directed downwards” and $F_\varepsilon$ the noise “directed upwards”, and hence their composition should be able to eliminate all the bumps. These operators are GENEOs with respect to isometries of the real line. We prove that moving in the space of GENEOs by taking suitable values for $\varepsilon$ and $\delta$, we can get quite close to restoring the original function $\varphi$, depending on how the bumps are positioned. The closer the bumps are to being Dirac delta functions and the further they are from each other, the better our approximation can be. On the ground of this result, we finally get an estimate of the expected value $E(\|F^\delta \circ F_\varepsilon(\hat{\varphi}) - \varphi\|_\infty)$.

The paper is structured as follows. In Section 2 the mathematical background is laid. The case we consider and our notation are explained in Section 3. In Section 4 we prove the results that are needed in order to demonstrate our main result. In Section 5 the main theorem giving us a probabilistic upper bound is formulated. Finally, in Section 6 some examples and experiments are presented in order to better illustrate the use of our results.
2 Mathematical setting

In this section we will recall some basic concepts we will use in this paper.

2.1 Representing data as real functions

Let us consider a set $\Phi$ of bounded functions from a set $X$ to $\mathbb{R}$, which will represent the data we wish to compare. We shall call $\Phi$ the set of admissible measurements on $X$. We endow $\Phi$ with the topology induced by the sup-norm $\| \cdot \|_\infty$ and the corresponding distance $D_\Phi$. A pseudo-metric $D_X$ can be defined on $X$ by setting $D_X(x_1, x_2) = \sup_{\varphi \in \Phi} |\varphi(x_1) - \varphi(x_2)|$ for every $x_1, x_2 \in X$. We recall that a pseudo-metric on a set $X$ is a distance $d$ without the property $d(x, y) = 0 \implies x = y$. We will consider the topological space $(X, \tau_{D_X})$ where $\tau_{D_X}$ is the topology induced by $D_X$. A base for this topology is given by the open balls $\{ B(x, r) := \{ x' \in X : D_X(x, x') < r \} \}$ with $x \in X$, $r \in \mathbb{R}$. The choice of this topology makes every function in $\Phi$ a continuous functions. As shown in [2], this fact enables us to use persistence diagrams in the study of $\Phi$.

2.2 GENEOs as operators acting on data

We are interested in considering transformations of data. Let $\text{Homeo}_\Phi(X)$ be the set of $\Phi$-preserving homeomorphisms from $X$ to $X$ with respect to the topology $\tau_{D_X}$, meaning that every $g$ in $\text{Homeo}_\Phi(X)$ is a homeomorphism of $X$ such as both $\varphi \circ g$ and $\varphi \circ g^{-1}$ belong to $\Phi$ for every $\varphi$ in $\Phi$. Let $G$ be a subgroup of $\text{Homeo}_\Phi(X)$. $G$ represents the set of transformations on data for which we will require equivariance to be respected.

Under the previously stated assumptions we call the ordered pair $(\Phi, G)$ a perception pair. We can now introduce the concept of GENEO.

**Definition 1** Let $(\Phi, G)$ and $(\Psi, H)$ be perception pairs and assume that a homomorphism $T : G \to H$ is given. A function $F : \Phi \to \Psi$ is called a Group Equivariant Non-Expansive Operator (GENEO) from $(\Phi, G)$ to $(\Psi, H)$ with respect to $T$ if the following properties hold:

1. (Group Equivariance) $F(\varphi \circ g) = F(\varphi) \circ T(g)$ for every $\varphi \in \Phi, g \in G$;
2. (Non-Expansivity) $D_\Psi(F(\varphi_1), F(\varphi_2)) \leq D_\Phi(\varphi_1, \varphi_2)$ for every $\varphi_1, \varphi_2 \in \Phi$.

Let us now consider the set $\mathcal{F}_{\text{all}}$ of all GENEOs from $(\Phi, G)$ to $(\Psi, H)$ with respect to $T : G \to H$. The space $\mathcal{F}_{\text{all}}$ is endowed with the extended pseudo-metric $D_{\mathcal{F}_{\text{all}}}$, defined by setting $D_{\mathcal{F}_{\text{all}}}(F_1, F_2) = \sup_{\varphi \in \Phi} D_\Psi(F_1(\varphi), F_2(\varphi))$ for every $F_1, F_2 \in \mathcal{F}_{\text{all}}$. The word extended refers to the possibility that $D_{\mathcal{F}_{\text{all}}}$ takes an infinite value.

The following result can be proven [2]:

**Theorem 1** If $(\Phi, D_\Phi), (\Psi, D_\Psi)$ are compact and convex then the metric space $(\mathcal{F}_{\text{all}}, D_{\mathcal{F}_{\text{all}}})$ is compact and convex.
If a non-empty set \( F \subseteq \mathcal{F} \) is fixed, we can define the following pseudo-distance \( D_{F, \Phi} \) on \( \Phi \):

**Definition 2** For any \( \varphi_1, \varphi_2 \) in \( \Phi \) we set

\[
D_{F, \Phi}(\varphi_1, \varphi_2) = \sup_{F \in F} \| F(\varphi_1) - F(\varphi_2) \|_\infty.
\]

This pseudo-distance allows us to compare data by taking into account how agents operate on data. Notice how, if \( G \) becomes larger, the natural pseudo-distance \( d_G \) becomes harder to compute but this new pseudo-distance \( D_{F, \Phi} \) becomes easier to evaluate. In order to find a lower bound for \( D_{F, \Phi} \) it is useful to introduce the notion of persistence diagram.

### 2.3 Persistence Diagrams

We will now recall some basic definitions and results in persistent homology. The interested reader can find more details in [11].

Let us consider an ordered pair \((X, \varphi)\) where \(X\) is a topological space and \(\varphi : X \to \mathbb{R}\) is a continuous function. For any \(t \in \mathbb{R}\) we can set \(X_t := \varphi^{-1}([-\infty, t])\). If \(u < v\) the inclusion \(i_{u,v} : X_u \to X_v\) induces a homomorphism \(i_{u,v}^k : H_k(X_u) \to H_k(X_v)\) between the \(k\)th homology groups of \(X_u\) and \(X_v\). We can define the \(k\)th persistent homology group, with respect to \(\varphi\) and computed at the point \((u, v)\), as \(PH_k(u, v) := i_{u,v}^k(H_k(X_u))\). Moreover, we can define the \(k\)th persistent Betti numbers function \(r_k(u, v)\) as the rank of \(PH_k(u, v)\).

The \(k\)th persistent Betti numbers function can be represented by the \(k\)th persistence diagram. This diagram is defined as the multi-set of all the ordered pairs \((u_j, v_j)\), where \(u_j\) and \(v_j\) are the times of birth and death of the \(j\)th \(k\)-dimensional hole in \(X\), respectively. We call **time of birth** of a hole the first time at which the homology class appears, and **time of death** the first time at which the homology class merges with an older one. When a hole never dies, we set its time of death equal to \(\infty\). We also add to this set all points of the form \((w, w)\) for \(w \in \mathbb{R}\).

In Figure 1 the filtration of the set \(X := [0, \frac{3}{4}\pi]\) given by the function \(\varphi(x) = 2 \sin x\) is illustrated. In this example, the topology on \(X\) is the one defined by the space \(\Phi\) of admissible functions given by all functions \(a \sin(x+b)\) with \(a, b \in \mathbb{R}\). The reader can easily check that this topology coincides with the Euclidean topology. The persistence diagram in degree \(k = 0\) of the function \(\varphi\) is displayed in Figure 2.

### 2.4 Comparing Persistence Diagrams

Persistence diagrams can be efficiently compared by means of a suitable metric \(d_{\text{match}}\). In order to define it, we first define the pseudo-distance

\[
\delta((x, y), (x', y')) = \min\{\max\{|x - x'|, |y - y'|\}, \max\{|x - y|/2, |x' - y'|/2\}\}
\]
Noise reduction through GENEOs

Fig. 1 How the sublevel sets change with respect to the filtration induced by $\varphi$.

Fig. 2 Persistence diagram of the function $2 \sin x$ on $[0, \frac{3}{4}\pi]$. The point $(0, \infty)$ describes the existence of a connected component that is born at zero and never dies. The point $(\sqrt{2}, 2)$ claims that there is a connected component born at $\sqrt{2}$ that dies (merges with the other one) at 2. The trivial points on the diagonal $u = v$ are not displayed.

for all $(x, y), (x', y') \in \{(x, y) \in \mathbb{R} \text{ with } x \leq y\} \cup \{(x, \infty) \text{ with } x \in \mathbb{R}\}$ by agreeing that $\infty - y = \infty, y - \infty = -\infty$ for $y \neq \infty, \infty - \infty = 0, \infty / 2 = \infty, |\pm \infty| = \infty, \min\{\infty, c\} = c, \max\{\infty, c\} = \infty$.

If two persistence diagrams $D, D'$ are given, we can set

$$d_{\text{match}}(D, D') = \inf_{\sigma \in \Sigma} \sup_{P \in D} \delta(P, \sigma(P))$$

where $\Sigma$ represents the set of all bijections between the multisets $D, D'$.

For every degree $k$ we can now define a new pseudo-metric:

$$D_{F, \Phi}^{\text{match}}(\varphi_1, \varphi_2) = \sup_{F \in \mathcal{F}} d_{\text{match}}(D_{F(\varphi_1)}, D_{F(\varphi_2)})$$
where $D_{F(\varphi_1)}, D_{F(\varphi_2)}$ are the persistence diagrams at degree $k$ of the functions $F(\varphi_1), F(\varphi_2)$, respectively.

In this paper we will limit ourselves to considering data represented as functions from $\mathbb{R}$ to $\mathbb{R}$, and we recall that for this kind of data persistence diagrams are non-trivial only in degree $k = 0$ (i.e., when persistent homology is used to count connected components). For this reason, in the following we will always assume $k = 0$.

### 3 Our model

In this paper we will be mainly interested in the set Lip$_L$ of all $L$-Lipschitz functions from $\mathbb{R}$ to $\mathbb{R}$, for some fixed constant $L \in \mathbb{R}$, and in the set $C^0(\mathbb{R})$ of all functions from $\mathbb{R}$ to $\mathbb{R}$ that are continuous with respect to the Euclidean topology. We will set $X = \mathbb{R}$ and consider the perception pairs (Lip$_L, G), (C^0(\mathbb{R}), G)$, where $G$ is the group of the Euclidean isometries of $\mathbb{R}$.

We will assume our noise to be made up of a finite number of copies of a “mother” nonnegative continuous bump function $\psi : \mathbb{R} \to \mathbb{R}$, such that $\text{supp}(\psi) \subseteq [-\sigma, \sigma]$ for some $\sigma > 0$ and $\|\psi\|_\infty \leq 1$. We recall that the support of a function is the closure of the set of points where $f$ is non-vanishing. After fixing two positive real numbers $\eta, \beta$, the noise we will be adding is a function $R$ belonging to the space $\mathcal{R}_{\eta,\beta}$ that contains the null function and all functions of the form $\sum_{i=1}^k a_i \psi(b_i(x - c_i))$, where $k$ is a positive integer and $a_i, b_i, c_i$ are real numbers such that $|c_i - c_j| \geq \eta$ for $i \neq j$ and $b_i \geq \beta$ for every index $i$. 

![Fig. 3](image-url) How drastically impulsive noise can influence persistence diagrams.
Noise reduction through GENEOs

Fig. 4 In the figure on the left we have our original function $\varphi$, in the middle a noise function $R$ and in the right figure the corrupted function $\hat{\varphi} := \varphi + R$.

For any $R \in \mathcal{R}_{\eta, \beta}$ we define $S(R) := \bigcup_{i=1}^{k} c_i - \frac{\sigma}{\beta}, c_i + \frac{\sigma}{\beta}$ and remark that if $x \not\in S(R)$ then $R(x) = 0$.

Our purpose will be the one of recovering $\varphi \in \text{Lip}_L$ as well as possible from the function $\hat{\varphi} = \varphi + R$. An example of such situation is depicted in Figure 4.

The following result will be of use.

**Proposition 2** Let $F_1, F_2$ be GENEOs from $(C^0(\mathbb{R}), G)$ to $(C^0(\mathbb{R}), G)$ with respect to the trivial homomorphism $T = \text{id} : G \to G$. Then $F_1 \circ F_2$ is a GENEO from $(C^0(\mathbb{R}), G)$ to $(C^0(\mathbb{R}), G)$ with respect to $T$.

**Proof** For every $\varphi \in C^0(\mathbb{R}), g \in G$ we have that

\[
F_1 \circ F_2(\varphi \circ g) = F_1(F_2(\varphi \circ g)) \\
= F_1(F_2(\varphi)) \circ g \\
= F_1(F_2(\varphi)) \circ g \\
= (F_1 \circ F_2)(\varphi) \circ g.
\]

Therefore, $F_1 \circ F_2$ is $G$-equivariant. Moreover, for any $\varphi_1, \varphi_2 \in C^0(\mathbb{R})$

\[
D_{C^0(\mathbb{R})}(F_1 \circ F_2(\varphi_1), F_1 \circ F_2(\varphi_2)) = D_{C^0(\mathbb{R})}(F_1(F_2(\varphi_1)), F_1(F_2(\varphi_2))) \\
\leq D_{C^0(\mathbb{R})}(F_2(\varphi_1), F_2(\varphi_2)) \\
\leq D_{C^0(\mathbb{R})}(\varphi_1, \varphi_2).
\]

It follows that $F_1 \circ F_2$ is non-expansive. \qed

4 Cutting off the noise by GENEOs

We start by introducing two families of GENEOs from $(C^0(\mathbb{R}), G)$ to $(C^0(\mathbb{R}), G)$ with respect to the identical homomorphism.

**Definition 3** Let $\varphi \in \text{Lip}_L$ and $\varepsilon > 0$. For all $x \in \mathbb{R}$ we define:

1. $F^\varepsilon(\varphi)(x) = \max(\varphi(x - \varepsilon), \varphi(x + \varepsilon))$;
2. $F_\varepsilon(\varphi)(x) = \min(\varphi(x - \varepsilon), \varphi(x + \varepsilon))$. 
Proposition 3 The maps $F^\varepsilon$ and $F_\varepsilon$ are GENEOs from $(C^0(\mathbb{R}), G)$ to $(C^0(\mathbb{R}), G)$ with respect to the identical homomorphism.

Proof We start by proving that $F^\varepsilon$ is $G$-equivariant.

Let $g$ be the translation $x \mapsto x + k$ with $k \in \mathbb{R}$, then

$$F^\varepsilon(\varphi \circ g) = \max\{\varphi((x + k) - \varepsilon), \varphi((x + k) + \varepsilon)\}$$

$$= \max\{\varphi((x - \varepsilon) + k), \varphi((x + \varepsilon) + k)\}$$

$$= F^\varepsilon(\varphi) \circ g.$$

Let $g$ be the symmetry $x \mapsto -x$, then

$$F^\varepsilon(\varphi \circ g) = \max\{\varphi((-x) - \varepsilon), \varphi((-x) + \varepsilon)\}$$

$$= \max\{\varphi(-(x + \varepsilon)), \varphi(-(x - \varepsilon))\}$$

$$= \max\{\varphi(x + \varepsilon), \varphi(x - \varepsilon)\} \circ g$$

$$= F^\varepsilon(\varphi) \circ g.$$

Since every isometry in $G$ can be written as the composition of a symmetry and a translation, our statement follows.

Furthermore, for any $x \in \mathbb{R}$

$$|F^\varepsilon(\varphi_1)(x) - F^\varepsilon(\varphi_2)(x)| =$$

$$= |\max\{\varphi_1(x - \varepsilon), \varphi_1(x + \varepsilon)\} - \max\{\varphi_2(x - \varepsilon), \varphi_2(x + \varepsilon)\}|$$

$$\leq \max\{|\varphi_1(x - \varepsilon) - \varphi_2(x - \varepsilon)|, |\varphi_1(x + \varepsilon) - \varphi_2(x + \varepsilon)|\}$$

$$\leq \max\{|\varphi_1 - \varphi_2|_{\infty}, |\varphi_1 - \varphi_2|_{\infty}\}$$

$$= |\varphi_1 - \varphi_2|_{\infty}.$$

It follows that $\|F^\varepsilon(\varphi_1) - F^\varepsilon(\varphi_2)\|_{\infty} \leq |\varphi_1 - \varphi_2|_{\infty}$, and hence $F^\varepsilon$ is non-expansive.

We observe that $\min\{a, b\} = -\max\{-a, -b\}$, and hence $F_\varepsilon(\varphi) = -F^\varepsilon(-\varphi)$ for every $\varphi \in C^0(\mathbb{R})$. Moreover, the map taking $\varphi$ to $-\varphi$ is a GENEO from $(C^0(\mathbb{R}), G)$ to $(C^0(\mathbb{R}), G)$ with respect to the identical homomorphism. Therefore, Proposition 2 guarantees that $G$-equivariance and non-expansivity also hold for $F_\varepsilon$. \hfill \Box

Since Proposition 2 shows that the composition of GENEOs is still a GENEO, the operator $F^\delta \circ F_\varepsilon$ is a GENEO from $(C^0(\mathbb{R}), G)$ to $(C^0(\mathbb{R}), G)$ with respect to the identical homomorphism.

We want now to prove that if a function $\hat{\varphi}$ is obtained by adding impulsive noise to a function $\varphi$ then the value of

$$\|F^\delta \circ F_\varepsilon(\hat{\varphi}) - \varphi\|_{\infty}$$

is bounded, and possibly small, provided that $\delta$ and $\varepsilon$ are suitably chosen. The main idea is that the operator $F_\varepsilon$ cuts the noise “directed upwards” and $F^\delta$ cuts the noise “directed downwards”.

In order to proceed, we need two lemmas.

Lemma 1 Let $R \in C^0(\mathbb{R}, \mathbb{R})$, $\varphi \in \text{Lip}_L$ for some $L \in \mathbb{R}$, and set $\hat{\varphi} := \varphi + R$. Then for any $\varepsilon > 0$ and $\delta > 0$
\[ \begin{align*}
&i) \quad -L\varepsilon + F_\varepsilon(R) \leq F_\varepsilon(\tilde{\varphi}) - \varphi \leq L\varepsilon + F_\varepsilon(R); \\
&ii) \quad -L\delta + F^\delta(R) \leq F^\delta(\tilde{\varphi}) - \varphi \leq L\delta + F^\delta(R); \\
&iii) \quad -L(\delta + \varepsilon) + F^\delta \circ F_\varepsilon(R) \leq F^\delta \circ F_\varepsilon(\tilde{\varphi}) - \varphi \leq L(\delta + \varepsilon) + F^\delta \circ F_\varepsilon(R).
\end{align*} \]

**Proof** Since \( \varphi \) is Lipschitz of constant \( L \) we have that for any value \( x \in \mathbb{R} \) \( |\varphi(x - \varepsilon) - \varphi(x)| \leq L\varepsilon \) and \( |\varphi(x + \varepsilon) - \varphi(x)| \leq L\varepsilon \). Therefore,

\[
F_\varepsilon(\tilde{\varphi})(x) = F_\varepsilon(\varphi + R)(x) = \min\{\varphi(x - \varepsilon) + R(x - \varepsilon), \varphi(x + \varepsilon) + R(x + \varepsilon)\} \leq \min\{\varphi(x) + L\varepsilon + R(x - \varepsilon), \varphi(x) + L\varepsilon + R(x + \varepsilon)\} = \varphi(x) + L\varepsilon + \min\{R(x - \varepsilon), R(x + \varepsilon)\} = \varphi(x) + L\varepsilon + F_\varepsilon(R)(x).
\]

Analogously, \( F^\delta \circ F_\varepsilon(\tilde{\varphi})(x) \geq \varphi(x) - L\varepsilon + F^\delta \circ F_\varepsilon(R)(x) \). The same steps applied to \( F^\delta \) yield the second statement of the lemma. As for the last claim we can see that:

\[
F^\delta \circ F_\varepsilon(\tilde{\varphi})(x) = \max\{F_\varepsilon(\tilde{\varphi})(x - \delta), F_\varepsilon(\tilde{\varphi})(x + \delta)\} \leq \max\{\varphi(x - \delta) + L\varepsilon + F_\varepsilon(R)(x - \delta), \varphi(x + \delta) + L\varepsilon + F_\varepsilon(R)(x + \delta)\} \leq \max\{\varphi(x) + L\delta + L\varepsilon + F_\varepsilon(R)(x - \delta), \varphi(x) + L\delta + L\varepsilon + F_\varepsilon(R)(x + \delta)\} = \varphi(x) + L\delta + L\varepsilon + \max\{F_\varepsilon(R)(x - \delta), F_\varepsilon(R)(x + \delta)\} = \varphi(x) + L\delta + L\varepsilon + F^\delta \circ F_\varepsilon(R)(x).
\]

Analogously, we can prove the lower bound. \( \square \)

Henceforth we will assume that any summation on an empty set of indexes is the null function.

**Lemma 2** Let \( R \in \mathcal{R}_{\eta,\beta} \) and \( \lambda \geq \frac{\eta}{2} \). If \( \lambda \leq \rho \leq \frac{\eta}{2} - \lambda \) then

a) \( F^\rho(R)(x) = \sum_{a_i < 0} [a_i \psi(b_i(x - \rho - c_i)) + a_i \psi(b_i(x + \rho - c_i))] \leq 0 \)

b) \( F^\rho(R)(x) = \sum_{a_i > 0} [a_i \psi(b_i(x - \rho + c_i)) + a_i \psi(b_i(x + \rho + c_i))] \geq 0 \)

for all \( x \in \mathbb{R} \). Moreover \( F^\rho(R), F^\rho(R)(x) \in \mathcal{R}_{2\lambda,\beta} \).

**Proof** We will suppose without loss of generality that \( c_i < c_{i+1} \) for all \( i = 1, \ldots, k - 1 \) and \( a_i \neq 0 \) for all indexes \( i \).

We want to show that at least one of \( x - \rho \) and \( x + \rho \) must always be outside of \( \bigcup_{\text{i}=1}^{k} |c_i - \lambda, c_i + \lambda| \). If by contradiction both \( x - \rho \) and \( x + \rho \) were in the same \( |c_i - \lambda, c_i + \lambda| \) then we would have \( \rho < \lambda \), against our hypotheses. Moreover \( x - \rho \) and \( x + \rho \) cannot belong to different intervals \( |c_i - \lambda, c_i + \lambda| \) and \( |c_j - \lambda, c_j + \lambda| \) for \( i < j \); if by contradiction \( x - \rho \in |c_i - \lambda, c_i + \lambda| \) and \( x + \rho \in |c_j - \lambda, c_j + \lambda| \) for some \( i < j \) then \( 2\rho = x + \rho - (x - \rho) > (c_j - \lambda) - (c_i + \lambda) \geq c_j - c_i - 2\lambda \geq \eta - 2\lambda \), and hence we would have \( \rho > \frac{\eta}{2} - \lambda \), against our hypotheses.
Since \( c_i - \frac{\sigma}{\beta}, c_i + \frac{\sigma}{\beta} \subseteq |c_i - \lambda, c_i + \lambda| \), it follows that at least one among the values \( R(x - \rho) \) and \( R(x + \rho) \) must always be zero. Let us now set \( I_i^- := |(c_i - \rho) - \lambda, (c_i - \rho) + \lambda| \) and \( I_i^+ := |(c_i + \rho) - \lambda, (c_i + \rho) + \lambda| \). These two intervals must be disjoint, since \( (c_i + \rho - \lambda) - (c_i - \rho + \lambda) = 2\rho - 2\lambda \geq 0 \).

Let us now consider \( \{I_1^-, I_1^+, \ldots, I_k^-, I_k^+\} \). We will now prove that any two distinct elements from this set must be disjoint. Since we have just proven that \( I_i^- \cap I_j^+ = \emptyset \) for \( i = 1, \ldots, k \), the following holds in the case \( i < j \):

1. \( I_i^+ \cap I_j^- = \emptyset \), since \( c_j + \rho - \lambda - (c_i + \rho + \lambda) \geq c_j - c_i - 2\lambda \geq \eta - 2\lambda \geq 0 \);
2. \( I_i^- \cap I_j^- = \emptyset \), since \( c_j - \rho - \lambda - (c_i - \rho + \lambda) \geq c_j - c_i - 2\lambda \geq \eta - 2\lambda \geq 0 \);
3. \( I_i^+ \cap I_j^- = \emptyset \), since \( c_j + \rho - \lambda - (c_i + \rho + \lambda) \geq c_j - c_i - 2\rho - 2\lambda \geq \eta - 2(\frac{\rho}{2} - \lambda) - 2\lambda = 0 \);
4. \( I_j^+ \cap I_i^- \), since \( c_j + \rho - \lambda - (c_i - \rho + \lambda) \geq c_j - c_i + 2\rho - 2\lambda \geq \eta + 2\lambda - 2\lambda \geq 0 \).

Let us now fix \( k \in \{1, \ldots, k\} \).

Since \( \text{supp}(a_k\psi(b_k(x - c_k))) \subseteq |c_k - \lambda, c_k + \lambda| \), then \( \text{supp}(a_k\psi(b_k(x - \rho - c_k))) \subseteq I_k^- \) and \( \text{supp}(a_k\psi(b_k(x + \rho - c_k))) \subseteq I_k^+ \).

Hence, for any \( x \in \mathbb{R} \) we have that
\[
F_{\rho}(R)(x) = \min\{R(x - \rho), R(x + \rho)\} = \begin{cases} 
0 & \text{if } x \notin \bigcup_{i=1}^{k} (I_i^- \cup I_i^+) \\
 a_j\psi(b_j(x + \rho - c_j)) & \text{if } x \in I_j^- \text{ and } a_j < 0 \\
 a_j\psi(b_j(x - \rho - c_j)) & \text{if } x \in I_j^+ \text{ and } a_j < 0.
\end{cases}
\]

This means that \( F_{\rho}(R)(x) = \sum_{a_i < 0} \left[a_i\psi(b_i(x - \rho - c_i)) + a_i\psi(b_i(x + \rho - c_i))\right] \leq 0 \).

Now, given \( c_i + \rho \) and \( c_j - \rho \) centers of bumps of \( F_{\rho}(R) \), we have that
\[ |(c_i + \rho) - (c_j - \rho)| \geq \min\{\eta - 2\rho, 2\rho\} \geq 2\lambda. \]

It follows that \( F_{\rho}(R) \in \mathcal{R}_{2\lambda, \beta} \).

By noticing that \( F^\delta(-R) = -F_{\rho}(-R) \) we get the second part of the thesis.

Let us remark that, in particular, if \( \lambda = 2\frac{\sigma}{\beta} \) then \( F_{\rho}(R), F^\rho(R) \in \mathcal{R}_{4\frac{\sigma}{\beta}, \beta} \).

We observe that in order to exist a \( \rho \) that satisfies the hypotheses of Lemma 2, the inequality \( \eta \geq 4\lambda \) must hold.

We are now actually ready to prove a key result in our paper.

**Theorem 4** Given \( \theta, \beta, L \in \mathbb{R} \), let \( \bar{R} \in \mathcal{R}_{\theta, \beta}, \varphi \in \text{Lip}_L \) and set \( \hat{\varphi} := \varphi + \bar{R} \). If \( 2\frac{\sigma}{\beta} \leq \varepsilon \leq \frac{\theta}{2} - 2\frac{\sigma}{\beta} \) then for any \( \frac{\sigma}{\beta} \leq \delta \leq \frac{1}{2} \min\{\theta - 2\varepsilon, 2\varepsilon\} - \frac{\sigma}{\beta} \) the following inequality holds:
\[
\|F^\delta \circ F_{\varepsilon}(\hat{\varphi}) - \varphi\|_{\infty} \leq L(\varepsilon + \delta).
\]

**Proof** Let \( x \in \mathbb{R} \). Firstly, let us notice that the condition \( 2\frac{\sigma}{\beta} \leq \varepsilon \leq \frac{\theta}{2} - 2\frac{\sigma}{\beta} \) implies that \( \frac{\sigma}{\beta} \leq \frac{1}{2} \min\{\theta - 2\varepsilon, 2\varepsilon\} - \frac{\sigma}{\beta} \). From Lemma 1 we know that
\[
-L(\delta + \varepsilon) + F^\delta \circ F_{\varepsilon}(\bar{R})(x) \leq F^\delta \circ F_{\varepsilon}(\hat{\varphi})(x) - \varphi(x) \leq L(\delta + \varepsilon) + F^\delta \circ F_{\varepsilon}(\bar{R})(x).
\]
Let us now prove that $F^\delta \circ F_\varepsilon(\bar{R})(x) = 0$.

By applying Lemma 2 with $\eta := \theta$, $\lambda := 2\frac{\sigma}{\beta}$, $\rho := \varepsilon$ and $R := \bar{R}$ we get

$$F_\varepsilon(\bar{R})(x) = \sum_{a_i < 0} [a_i \psi(b_i(x - \varepsilon - c_i)) + a_i \psi(b_i(x + \varepsilon - c_i))] \leq 0.$$  

Let us remark that $F_\varepsilon(\bar{R}) \in \mathcal{R}\frac{\theta}{\beta}$, $\beta$. Moreover,

1. $F^\delta \circ F_\varepsilon(\bar{R})(x) = \max\{F_\varepsilon(\bar{R})(x - \delta), F_\varepsilon(\bar{R})(x + \delta)\} \leq 0$ since both terms are negative;
2. $F^\delta \circ F_\varepsilon(\bar{R})(x) = F^\delta(F_\varepsilon(\bar{R}))(x) \geq 0$.

These inequalities follow from Lemma 2, by setting $\eta := 4\frac{\sigma}{\beta}$, $\lambda := \frac{\sigma}{\beta}$, $\rho := \delta$ and $R := F_\varepsilon(\bar{R})$.

Therefore, we have proved that $|F^\delta \circ F_\varepsilon(\hat{\varphi})(x) - \varphi(x)| \leq L(\delta + \varepsilon)$ for any $x \in \mathbb{R}$.

It follows that $\|F^\delta \circ F_\varepsilon(\hat{\varphi}) - \varphi\|_\infty \leq L(\varepsilon + \delta)$. \hfill \Box

Let us remark that Theorem 4 works under the (only) implicit assumption that $\theta \geq 8\frac{\sigma}{\beta}$. In our setting this should not be restrictive since it means that the noise added is made up of scattered, thin bumps, without any reference to the height of the bumps: this is what we expect when considering additive impulsive noise.

**Corollary 5** Given $\theta, \beta, L \in \mathbb{R}$, let $R \in \mathcal{R}\theta, \beta$, $\varphi \in \text{Lip}_L$, and set $\hat{\varphi} := \varphi + R$. If $\theta \geq 8\frac{\sigma}{\beta}$, then

$$\|F^{\frac{\sigma}{\beta}} \circ F^{2\frac{\sigma}{\beta}}(\hat{\varphi}) - \varphi\|_\infty \leq 3L\frac{\sigma}{\beta}.$$

**Proof** The claim follows from Theorem 4 by taking $\varepsilon := 2\frac{\sigma}{\beta}$ and $\delta := \frac{\sigma}{\beta}$. \hfill \Box

Corollary 5 and the well known stability of persistence diagrams with respect to the max-norm [8] immediately imply the following result, which is of interest in TDA (the symbol $d_{\text{match}}$ denotes the usual bottleneck distance between persistence diagrams).

**Corollary 6** Given $\theta, \beta, L \in \mathbb{R}$, let $R \in \mathcal{R}\theta, \beta$, $\varphi \in \text{Lip}_L$, and set $\hat{\varphi} := \varphi + R$. If $\theta \geq 8\frac{\sigma}{\beta}$, and $D$ and $D'$ are the persistence diagrams in degree 0 of the filtering functions $\varphi$ and $F^{\frac{\sigma}{\beta}} \circ F^{2\frac{\sigma}{\beta}}(\hat{\varphi})$, respectively, then

$$d_{\text{match}}(D, D') \leq 3L\frac{\sigma}{\beta}.$$  

5 Our main result

We are now ready to prove our main result. We start by stating a lemma concerning the probability $p$ that any two distinct points in a randomly chosen set of cardinality $k$ in an interval of length $\ell$ have a distance greater than $\eta$. A
proof of the following lemma is provided in \cite{14}: for the reader’s convenience, we report it here.

**Lemma 3** Let $X_1, \ldots, X_k$, with $k \geq 2$, be independent random variables, uniformly distributed on the interval $[0, \ell]$, for some $\ell > 0$. Let

$$M := \min_{i \neq j} |X_i - X_j|$$

be the minimal distance between two distinct random variables. Then we have

$$P(M > \eta) = \begin{cases} 
1 & \text{if } \eta \leq 0, \\
\left(1 - \frac{(k-1)\eta}{\ell}\right)^k & \text{if } 0 < \eta < \frac{\ell}{k-1}, \\
0 & \text{if } \eta \geq \frac{\ell}{k-1}.
\end{cases}$$

**Proof** It suffices to consider the case $0 < \eta < \frac{\ell}{k-1}$. By symmetry, we have

$$P(M > \eta) = k! \frac{\text{Leb}(S)}{\ell^k}$$

where $S = \{x \in [0, \ell]^k \mid x_1 < x_2 - \eta < x_3 - 2\eta < \cdots < x_k - (k-1)\eta\}$ and $\text{Leb}$ denotes the Lebesgue measure. Setting $y_i = x_i - (i-1)\eta$ for $i = 1, \ldots, k$, we have

$$S' = \{y \in [0, \ell - (k-1)\eta]^k \mid y_1 < y_2 < \cdots < y_k\}.$$ 

On the other hand, again by symmetry, we have

$$\text{Leb}(S') = \frac{\text{Leb}([0, \ell - (k-1)\eta]^k)}{k!} = \frac{(\ell - (k-1)\eta)^k}{k!}$$

and plugging this last identity into (1) we get the thesis. \(\square\)

We can now prove the following result, concerning the expected value of the error $\|F_{\sigma} \circ F_{2\beta}(\hat{\varphi}) - \varphi\|_\infty$.

**Theorem 7** Let us choose a function $\varphi \in \text{Lip}_L$, a non-negative continuous function $\psi : \mathbb{R} \to \mathbb{R}$ with $\|\psi\|_\infty \leq 1$ and $\text{supp}(\psi) \subseteq [-\sigma, \sigma]$ for some $\sigma > 0$, two positive numbers $\beta$ and $\ell$, and an integer $k \geq 2$. For $i = 1, \ldots, k$, let us fix $a_i \in \mathbb{R}$ and $b_i \geq \beta$, and set $\bar{\alpha} := \max |a_i|$. Moreover, let $c_1, \ldots, c_k$ be independent random variables, uniformly distributed on the interval $[0, \ell]$. Let us consider the random variable $\hat{\varphi} := \varphi + R$, where $R(x) := \sum_{i=1}^k a_i \psi(b_i(x - c_i))$ for any $x \in \mathbb{R}$. If $\frac{\sigma}{\beta} < \frac{\ell}{8(k-1)}$, then

$$E \left( \|F_{\sigma} \circ F_{2\beta}(\hat{\varphi}) - \varphi\|_\infty \right) \leq 3L \frac{\sigma}{\beta} + k\bar{\alpha} \left(1 - \left(1 - 8\frac{(k-1)\sigma}{\beta}\right)^k\right).$$
14 Noise reduction through GENEOs

Proof By setting $\delta = \sigma_\beta$ and $\varepsilon = 2\sigma_\beta$ in statement iii) of Lemma 1, we have that
\[
\|F_\sigma^{\beta} \circ F_{2\sigma}^{\beta} (\hat{\phi}) - \varphi\|_\infty \leq 3L\sigma_\beta + \|F_\sigma^{\beta} \circ F_{2\sigma}^{\beta} (R)\|_\infty.
\]
Since the operator $F_\sigma^{\beta} \circ F_{2\sigma}^{\beta}$ is non-expansive and $F_\sigma^{\beta} \circ F_{2\sigma}^{\beta} (0) = 0$, it follows that
\[
\|F_\sigma^{\beta} \circ F_{2\sigma}^{\beta} (R)\|_\infty \leq \|R\|_\infty \leq k\bar{\alpha} \|\psi\|_\infty \leq k\bar{\alpha}.
\]
Therefore, $\|F_\sigma^{\beta} \circ F_{2\sigma}^{\beta} (\hat{\phi}) - \varphi\|_\infty \leq 3L\sigma_\beta + k\bar{\alpha}$. If we apply Lemma 3 with $\eta = 8\sigma_\beta$, we obtain that $R \in R_{8\sigma_\beta, \beta}$ with probability $p := (1 - 8(k-1)\sigma_\beta^k)$. If $R \in R_{8\sigma_\beta, \beta}$, we can apply Theorem 4 by setting $\delta = \sigma_\beta$, $\varepsilon = 2\sigma_\beta$ and $\theta = 8\sigma_\beta$, and hence we obtain that $\|F_\sigma^{\beta} \circ F_{2\sigma}^{\beta} (\hat{\phi}) - \varphi\|_\infty \leq 3L\sigma_\beta$ with probability at least $p$. Since $\|F_\sigma^{\beta} \circ F_{2\sigma}^{\beta} (\hat{\phi}) - \varphi\|_\infty \leq 3L\sigma_\beta + k\bar{\alpha}$ in any case, it follows that
\[
E \left( \|F_\sigma^{\beta} \circ F_{2\sigma}^{\beta} (\hat{\phi}) - \varphi\|_\infty \right) \leq 3L\sigma_\beta p + \left( 3L\sigma_\beta + k\bar{\alpha} \right) (1 - p) = 3L\sigma_\beta + k\bar{\alpha} (1 - p).
\]
\[
\square
\]

Theorem 7 and the well known stability of persistence diagrams with respect to the max-norm \cite{8} immediately imply the following result, which is of interest in TDA (the symbol $d_{\text{match}}$ denotes the usual bottleneck distance between persistence diagrams).

Corollary 8 Let us make the same assumptions of Theorem 7. Let $D$ and $D'$ be the persistence diagrams in degree 0 of the filtering functions $\varphi$ and $F_\sigma^{\beta} \circ F_{2\sigma}^{\beta} (\hat{\phi})$, respectively. If $\frac{\sigma_\beta}{\beta} < \frac{\ell}{8(k-1)}$, then
\[
E \left( d_{\text{match}} (D, D') \right) \leq 3L\sigma_\beta + k\bar{\alpha} \left( 1 - \left( 1 - 8 \left( \frac{k-1}{\ell} \right) \sigma_\beta^k \right) \right).
\]

This result shows that the use of suitable GENEOs can make TDA (relatively) stable also in the presence of impulsive noise, under the assumptions we have considered in this paper.

6 Examples and experiments

We will now validate our approach based on GENEOs by giving two examples and illustrating some experimental results.

6.1 Examples

In order to verify how our approach works, we will set $\tau_n := (1 - \frac{1}{n}) 2\sigma_\beta + \frac{1}{n} \left( \frac{\theta}{2} - 2\sigma_\beta \right)$ and consider the upper bound $\|F_\sigma^{\tau_n} \circ F_{\tau_n} (\hat{\phi}) - \varphi\|_\infty \leq \frac{3}{2} L\tau_n$,
obtained by applying Corollary 5. We observe that \( \tau_n \geq 2\sigma_\beta \) for every index \( n \), and \( \lim_{n \to +\infty} \tau_n = 2\sigma_\beta \). We will examine two examples that use the GENEOs \( F_{\tau_n}^\varphi \circ F_{\tau_n} \), and show how our method based on such operators and the method based on convolutions differ, as for their capability in removing additive impulsive noise. Moreover, we will compare the actual error \( \| F_{\tau_n}^\varphi \circ F_{\tau_n} (\hat{\varphi}) - \varphi \|_\infty \) to its upper bound \( \frac{3}{2} L \tau_n \), by running several simulations. The convolutions that will be applied in our examples use the functions \( T_h : \mathbb{R} \to \mathbb{R} \) defined by setting

\[
T_h(x) := \begin{cases} 
\frac{h}{2} & \text{if } -\frac{1}{h} \leq x \leq \frac{1}{h} \\
0 & \text{otherwise}
\end{cases}
\]

for \( h > 0 \). We will see that, although the convolution with such functions is also a GENEO, it will not be able to efficiently remove the noise.

Our noise function \( R \) will be constructed starting from the mother function \( \psi \) defined by setting

\[
\psi(x) := e^{1 - \frac{1}{1-x^2}} \quad \text{for } x \in ]-1,1[ \quad \text{and } \psi(x) := 0 \quad \text{for } x \notin ]-1,1[.
\]

Using the notation introduced in the previous sections, we will set \( \sigma = 1.1 \), thus satisfying the condition \( \text{supp}(\psi) \subseteq [-\sigma, \sigma] \). The impulsive noise will be added in an interval \([-\ell, \ell]\). The following parameters are considered, with these respective uniform distributions:

- \( k \sim \text{Unif}\{1,...,10\} \)
- \( a_i \sim \text{Unif}\{-100,100\} \) for \( i = 1, \ldots, k \)
- \( b_i \sim \text{Unif}\{0,100\} \) for \( i = 1, \ldots, k \)
- \( c_i \sim \text{Unif}\{-\ell + \frac{\sigma}{\beta}, \ell - \frac{\sigma}{\beta}\} \) for \( i = 1, \ldots, k \).

We set \( \beta := \min_{i=1,...,N} b_i, \bar{\alpha} := \max_{i=1,...,N} |a_i|, \) and \( \eta := \min_{i \neq j} |c_i - c_j| \). After producing random values for the parameters \( N, a_i, b_i, c_i \), our algorithm checks whether \( \eta > 8\sigma_\beta \), otherwise it generates another set of parameters.

### 6.1.1 First example

Let us consider the function \( \varphi(x) := \begin{cases} 
\sin x & \text{if } -4\pi \leq x \leq 4\pi \\
0 & \text{otherwise}
\end{cases} \)

for \( x \in \mathbb{R} \). We observe that \( \varphi \in \text{Lip}_L \), for \( L = 1 \). We will add noise in the interval \([-4\pi,4\pi]\) (meaning \( \ell = 4\pi \)) and visualize the results in such an interval. Figure 5 illustrates how the function \( \hat{\varphi} \) looks like compared to \( \varphi \).

We will start by considering how well the convolution \( \hat{\varphi} \circ T_n \) can approximate the original function \( \varphi \), when \( n \) goes from 3 to 100. From Figure 6, it is immediately apparent that the max-norm distance between \( \hat{\varphi} \circ T_n \) and \( \varphi \) remains quite large. If we apply a convolution with \( T_\frac{\eta}{\beta} \), for \( 3 \leq n \leq 100 \), we get the results displayed in Figure 7, showing that all information represented by the function \( \varphi \) is progressively destroyed. In contrast, if we apply the operator \( F_{\tau_n}^\varphi \circ F_{\tau_n} \), for \( n = 3, 5, 20, 100 \), we get the results displayed...
Noise reduction through GENEOs

Fig. 5 Example 1: Comparison between $\varphi$ and $\hat{\varphi}$.

Fig. 6 Example 1: Denoising via convolution with $T_h$ ($h = n$).

In Figure 8. As we can see, this operator is much more efficient in removing the bumps and restoring the function $\varphi$.

As a matter of fact, when we apply a convolution with the function $T_h$ and check the corresponding errors via the sup-norm, we get the results displayed in Figure 9. Since $\lim_{h \to +\infty} \hat{\varphi} * T_h = \hat{\varphi}$ and $\lim_{h \to +\infty} \hat{\varphi} * \frac{T_1}{h} = 0$, we get that the errors tend to $\| \hat{\varphi} - \varphi \|_{\infty} = \max_{i=1, \ldots, N} |a_i| = \bar{\alpha}$ and $\| \varphi \|_{\infty}$, respectively.

In contrast, if we apply the operator $F_{\frac{2\sigma}{\beta}} \circ F_{\tau_n}$, we get the results displayed in Figure 10, showing that the upper bound for the error stated in Corollary 5 is quite tight. As we can expect, we get the best denoising by replacing $\tau_n$ with $2\frac{\sigma}{\bar{\beta}}$ (see Figure 11).

Finally, we executed 1000 simulations. In each of them we have produced a function $\hat{\varphi}$ by adding random impulsive noise to the function $\varphi$, then we have applied $F_{\frac{\sigma}{\beta}} \circ F_{2\frac{\sigma}{\bar{\beta}}}$ to $\hat{\varphi}$, in order to see how tight the upper bound in
Corollary 5 is to the actual error $\| F_{\hat{\varphi}} \circ F_{2\varphi}(\hat{\varphi}) - \varphi \|_\infty$. The same parameters as in the beginning of this example have been used. As we can see in Figure 12 the overestimation committed by our upper bound is often quite close to zero, relatively to the Lipschitz constant $L$ of the function $\varphi$. This suggests that such an upper bound is quite accurate.

### 6.1.2 Second example

Let us consider the function

$$\varphi(x) := \begin{cases} 
\frac{1}{1000}(x - 5)(x - 3)(x + 1)(x + 4)(x + 5) & \text{if } -5 \leq x \leq 5 \\
0 & \text{otherwise}
\end{cases}$$

for $x \in \mathbb{R}$. The coefficient $\frac{1}{1000}$ was chosen in order to get that the Lipschitz constant $L$ is comparable to the one of the previous example. In this example $L = \frac{27}{35}$. Figure 13 illustrates how the function $\hat{\varphi}$ looks like compared to $\varphi$.

We will start by considering how well the convolution $\hat{\varphi} \ast T_n$ can approximate the original function $\varphi$, when $n$ goes from 3 to 100. From Figure 14, it is immediately apparent that the max-norm distance between $\hat{\varphi} \ast T_n$ and $\varphi$ remains quite large.
Fig. 8 Example 1: Denoising via the proposed GENEo $F^{\frac{T_n}{2}} \circ F_{\tau_n}$.

Fig. 9 Example 1: Error made by applying a convolution with $T_n$ (left) or $T_{1/n}$ (right).

If we apply a convolution with $T_{1/n}$ for $3 \leq n \leq 100$, we get the results displayed in Figure 15, showing that all information represented by the function $\phi$ is progressively destroyed. In contrast, if we apply the operator $F^{\frac{T_n}{2}} \circ F_{\tau_n} (\hat{\phi})$, for $n = 3, 5, 20, 100$, we get the results displayed in Figure 16. As we can see, this operator is much more efficient in removing the bumps and restoring the function $\phi$.

When we apply a convolution with the function $T_h$ and check the corresponding errors via the sup-norm, we get the results displayed in Figure 17. As already seen, since $\lim_{h \to +\infty} \hat{\phi} * T_h = \hat{\phi}$ and $\lim_{h \to +\infty} \hat{\phi} * T_{1/n} = 0$ we get that the errors tend to $\|\hat{\phi} - \varphi\|_{\infty} = \max_{i=1,\ldots,N} |a_i| = \bar{\alpha}$ and $\|\varphi\|_{\infty}$, respectively.

In contrast, if we apply the operator $F^{\frac{T_n}{2}} \circ F_{\tau_n}$, we get the results displayed in Figure 18, showing that the upper bound for the error stated in Corollary 5
Fig. 10  Example 1: Error made by using the GENEO $F^{\tau_n} \circ F_{\tau_n}$.

Fig. 11  Example 1: Denoising via $F^{\sigma_2} \circ F_{2 \sigma_2}$.

Fig. 12  Example 1: Histogram counting the number of cases in each of ten bins, concerning the overestimation value $3L \sigma_\beta - \|F^{\sigma_\beta} \circ F_{2 \sigma_\beta}(\hat{\phi}) - \phi\|_\infty$ obtained in our simulations. Most of the cases belong to the first bin, containing the functions $\phi$ for which $0 \leq 3L \sigma_\beta - \|F^{\sigma_\beta} \circ F_{2 \sigma_\beta}(\hat{\phi}) - \phi\|_\infty \leq 0.02$.

is quite tight. As we can expect, we get the best denoising by replacing $\tau_n$ with $2 \sigma_\beta$ (see Figure 19).

Finally, we again executed 1000 simulations, using the same methodology as in the previous case, this time considering the polynomial presented at the beginning of this example. As we can see in Figure 20 the overestimation committed by our upper bound is often quite close to zero, relatively to the Lipschitz constant $L$ of the polynomial $\phi$. 
6.2 Experiments

In order to check how good the upper bound stated in Theorem 7 is, we have made the following experiment. We have set $\ell = 20$ and randomly generated a collection of 1000 Lipschitz functions. Each function has been generated as follows. First of all we have randomly chosen four values $k, \alpha, \beta, L$ with $k \in \{1, \ldots, 10\}$, $\alpha \in \{50, 55, 60, \ldots, 100\}$, $\beta \in \{0.5, 1, 1.5, \ldots, 5\}$, and $L \in \{1, 2, \ldots, 10\}$, with uniform distributions. Then we have randomly chosen and sorted in ascending order $k$ points $x_1, \ldots, x_k$ in the open interval $]0, \ell[$, with uniform distribution. Hence we have obtained the following decomposition: $]0, \ell[ = ]0, x_1[ \cup ]x_1, x_2[ \cup \ldots \cup ]x_k, \ell[$. We have defined our Lipschitz function $\varphi$ equal to 0 outside $]0, \ell[$. After setting $(x_0, y_0) = (0, 0)$ and $(x_{k+1}, y_{k+1}) = (\ell, 0)$, for $i \in \{1, \ldots, k + 1\}$ the value $y_i$ of the function at $x_i$ has been randomly chosen, with uniform distribution in an interval that allows for an $L$-Lipschitz extension to $[0, \ell]$ of the function, i.e.,
Finally, the graph of the Lipschitz function $\varphi$ on $[0, \ell]$ has been obtained by connecting each point $(x_{i-1}, y_{i-1})$ to $(x_i, y_i)$ with a segment, for $i \in \{1, \ldots, k+1\}$. We observe that $\varphi$ is a $L$-Lipschitz function. After that, we have considered the mother function $\psi$ defined by setting $\psi(x) := e^{1-\frac{1}{1-x^2}}$ for $x \in [-1, 1]$ and $\psi(x) := 0$ for $x \notin [-1, 1]$. Furthermore, we have fixed $\sigma = 1.1$.

For each $L$-Lipschitz function $\varphi$ produced in this way, we have considered the function $\hat{\varphi} = \varphi + \sum_{i=1}^{k} a_i \psi(b_i(x-c_i))$, where the parameters $a_i, b_i, c_i$ have been randomly chosen, according to the following uniform distributions:

- $a_i \sim \text{Unif}[0, \alpha]$ for $i = 1, \ldots, k$
- $b_i \sim \text{Unif}[\beta, 20]$ for $i = 1, \ldots, k$
- $c_i \sim \text{Unif}[3\sigma \beta, \ell-3\sigma \beta]$ for $i = 1, \ldots, k$.

In Figures 21 and 22 some examples of the functions we have produced are displayed, for $k = 3$ and $k = 7$, respectively. In each figure, the functions are displayed without noise (left) and with added noise (right).

After that, we have compared the probabilistic upper bound stated in Theorem 7 with the average of the value $\left\| F^\varphi \circ F_2^\varphi (\hat{\varphi} - \varphi) \right\|_\infty$, computed on our collection of $L$-Lipschitz functions for the chosen values of one of the three variables $\alpha, \beta, L$, when the other two vary. The results of this comparison are displayed in Figures 23, 24, and 25.

---

**Fig. 15** Example 2: Denoising via convolution with $T_h$ ($h = \frac{1}{n}$).
Noise reduction through GENEos

Fig. 16 Example 2: Denoising via the proposed GENEo $F^\tau_{\frac{2}{n}} \circ F_{\tau_n}$.

Acknowledgements

This research has been partially supported by INdAM-GNSAGA.

Conflict of interest

The authors declare that they have no conflict of interest.

Authors’ contributions

P.F. devised the project. All authors contributed to the manuscript. All authors read and approved the final manuscript. The authors of this paper have been listed in alphabetical order.
Fig. 18 Example 2: Error made by using the GENEO $F_{\tau n} \circ F_{\sigma}$. 

Fig. 19 Example 2: Denoising via $F_{\sigma} \circ F_{2 \sigma}$. 

Fig. 20 Example 2: Histogram counting the number of cases in each of ten bins, concerning the overestimation value $3L_{\sigma, \beta} - \|F_{\beta} \circ F_{2 \beta} (\hat{\varphi}) - \varphi\|_{\infty}$ obtained in our simulations. Most of the cases belong to the first two bins, containing the functions $\varphi$ for which $0 \leq 3L_{\sigma, \beta} - \|F_{\beta} \circ F_{2 \beta} (\hat{\varphi}) - \varphi\|_{\infty} \leq 0.24$.

References

[1] Robert J. Adler and Sarit Agami. Modelling persistence diagrams with planar point processes, and revealing topology with bagplots. *Journal of Applied and Computational Topology*, 3(3):139–183, Sep 2019.

[2] Mattia G. Bergomi, Patrizio Frosini, Daniela Giorgi, and Nicola Quercioli. Towards a topological–geometrical theory of group equivariant non-expansive operators for data analysis and machine learning. *Nature*
Noise reduction through GENEOs

Fig. 21 Five examples of the functions we have produced for $k = 3$. In each figure, the functions are displayed without noise (left) and with added noise (right).

Fig. 22 Five examples of the functions we have produced for $k = 7$. In each figure, the functions are displayed without noise (left) and with added noise (right).

Fig. 23 The difference between the probabilistic upper bound $3L\frac{\sigma}{\beta} + k\alpha \left(1 - \left(1 - 8\frac{(k-1)\sigma}{\ell}\right)^{k}\right)$ stated in Theorem 7 and the average of the value $\left\|F_{\frac{\sigma}{\beta}} \circ F_{\sigma,\beta}(\hat{\phi}) - \varphi\right\|_{\infty}$ with respect to varying $\beta$ and $L$, computed on our collection of $L$-Lipschitz functions for $\alpha$ varying in the set $\{50, 55, 60, \ldots, 100\}$.

Machine Intelligence, 1(9):423–433, Sep 2019.

[3] Silvia Biasotti, Leila De Floriani, Bianca Falcidieno, Patrizio Frosini, Daniela Giorgi, Claudia Landi, Laura Papaleo, and Michela Spagnuolo. Describing shapes by geometrical-topological properties of real functions. ACM Comput. Surv., 40(4):12:1–12:87, October 2008.
The difference between the probabilistic upper bound $3L \frac{\sigma}{\beta} + k\alpha \left(1 - \frac{1}{L} \left(1 - 8\left(\frac{k-1}{L}\right) \frac{\sigma}{\beta}\right)\right)^k$ stated in Theorem 7 and the average of the value $\|F^{\frac{\sigma}{\beta}} \circ F_{2\frac{\sigma}{\beta}} (\hat{\varphi}) - \varphi\|_{\infty}$ with respect to varying $\alpha$ and $L$, computed on our collection of $L$-Lipschitz functions for $\beta$ varying in the set $\{0.5, 1, 1.5, \ldots, 5\}$.

The difference between the probabilistic upper bound $3L \frac{\sigma}{\beta} + k\alpha \left(1 - \frac{1}{L} \left(1 - 8\left(\frac{k-1}{L}\right) \frac{\sigma}{\beta}\right)\right)^k$ stated in Theorem 7 and the average of the value $\|F^{\frac{\sigma}{\beta}} \circ F_{2\frac{\sigma}{\beta}} (\hat{\varphi}) - \varphi\|_{\infty}$ with respect to varying $\alpha$ and $\beta$, computed on our collection of $L$-Lipschitz functions for $L$ varying in the set $\{1, 2, \ldots, 10\}$.

[4] Mickaël Buchet, Frédéric Chazal, Tamal K. Dey, Fengtao Fan, Steve Y. Oudot, and Yusu Wang. Topological analysis of scalar fields with outliers. In Lars Arge and János Pach, editors, 31st International Symposium on Computational Geometry (SoCG 2015), volume 34 of Leibniz International Proceedings in Informatics (LIPIcs), pages 827–841, Dagstuhl, Germany, 2015. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[5] Gunnar Carlsson. Topology and data. Bull. Amer. Math. Soc. (N.S.), 46(2):255–308, 2009.

[6] Andrea Cerri, Barbara Di Fabio, Massimo Ferri, Patrizio Frosini, and Claudia Landi. Betti numbers in multidimensional persistent homology are stable functions. Math. Methods Appl. Sci., 36(12):1543–1557, 2013.

[7] Andrea Cerri, Marc Ethier, and Patrizio Frosini. On the geometrical properties of the coherent matching distance in 2D persistent homology.
Noise reduction through GENEOs

Journal of Applied and Computational Topology, 3(4):381–422, Dec 2019.

[8] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. Discrete Comput. Geom., 37(1):103–120, 2007.

[9] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko. Lipschitz functions have Lp-stable persistence. Foundations of Computational Mathematics, 10(2):127–139, Apr 2010.

[10] Francesco Conti, Patrizio Frosini, and Nicola Quercioli. On the construction of Group Equivariant Non-Expansive Operators via permutants and symmetric functions, Frontiers in Artificial Intelligence, 5:1–11, 2022.

[11] Herbert Edelsbrunner and John Harer. Persistent homology—a survey. In Surveys on discrete and computational geometry, volume 453 of Contemp. Math., pages 257–282. Amer. Math. Soc., Providence, RI, 2008.

[12] Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, and Aarti Singh. Confidence sets for persistence diagrams. The Annals of Statistics, 42(6):2301 – 2339, 2014.

[13] Patrizio Frosini and Grzegorz Jabłoński. Combining persistent homology and invariance groups for shape comparison. Discrete Comput. Geom., 55(2):373–409, 2016.

[14] Mike Earnest (https://math.stackexchange.com/users/177399/mike earnest). Average minimum distance between n points generate i.i.d. with uniform dist. Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/2001026 (version: 2016-11-09).

[15] Saeed V. Vaseghi. Impulsive Noise: Modelling, Detection and Removal, chapter 13, pages 341–358. John Wiley and Sons, Ltd, 2008.

[16] Siddharth Vishwanath, Kenji Fukumizu, Satoshi Kuriki, and Bharath K. Sriperumbudur. Robust persistence diagrams using reproducing kernels. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, Advances in Neural Information Processing Systems, volume 33, pages 21900–21911. Curran Associates, Inc., 2020.