## Communications in Optimization Theory

Available online at http://cot.mathres.org

# SUBSTITUTION AND COMPOSITION OPERATORS: THE EXPECTED AND UNEXPECTED 

## JÜRGEN APPELL

Department of Mathematics, University of Würzburg, Emil-Fischer-Str. 30, D-97074 Würzburg, Germany
Dedicated with infinite gratitude to the memory of Pjotr Petrovich Zabrejko, a distinguished scientist, wonderful teacher, and true friend


#### Abstract

We give continuity and compactness conditions for the linear substitution operator $S_{\varphi}(f)(t):=f(\varphi(t))$ and the nonlinear composition operator $C_{g}(f)(x):=g(f(x))$, in terms of the generating functions $\varphi:[0,1] \rightarrow[0,1]$ resp. $g: \mathbb{R} \rightarrow \mathbb{R}$, with a particular emphasis on examples and counterexamples. Keywords. Compactness; Composition operator; Continuity; Lipschitz continuity; Substitution operator. 2020 Mathematics Subject Classification. 47H30, 26A16, $26 A 45$. The author of this survey has written not less than 50 papers jointly with Pjotr Zabrejko. Up to 1990, many of these papers were dedicated to the study of superposition operators, a fact which led to the publication of the book [6]. He taught me what is mathematics, and we became true friends. Therefore I remember him with infinite gratitude, admiration, and appreciation, but also with profound sadness since he has been taken away from us in 2019, shortly after his 80th birthday which we celebrated together in Minsk.


## 1. Spaces and Operators

We start with describing the three spaces we will work in. Without loss of generality, we consider only real functions defined on the interval $[0,1]$. By $C$ we denote the linear space of all continuous functions $f:[0,1] \rightarrow \mathbb{R}$, equipped with the usual norm

$$
\|f\|_{C}:=\max \{|f(t)|: 0 \leq t \leq 1\} \quad(f \in C),
$$

by Lip the linear subspace of $C$ of all Lipschitz continuous functions $f:[0,1] \rightarrow \mathbb{R}$ with norm

$$
\|f\|_{L i p}:=|f(0)|+l i p(f) \quad(f \in L i p)
$$

where $\operatorname{lip}(f)$ denotes the minimal Lipschitz constant of $f$, and by $B V$ the linear space of all functions $f:[0,1] \rightarrow \mathbb{R}$ of bounded variation with norm

$$
\|f\|_{B V}:=|f(0)|+\operatorname{var}(f) \quad(f \in B V)
$$

E-mail address: jurgen@dmuw.de
Received: May 20, 2023; Accepted: July 16, 2023.
where $\operatorname{var}(f)$ denotes the Jordan variation of $f$. It is well-known that all these spaces are linear spaces and algebras, i.e., closed under addition and multiplication of functions, and also Banach spaces when equipped with the indicated norms. These spaces are so different that a parallel study of operators and equations in them becomes a rewarding and interesting task.

In the sequel, we study two operators in the above mentioned spaces. The first is the substitution operator (also called inner superposition operator)

$$
\begin{equation*}
S_{\varphi}(f)(t):=(f \circ \varphi)(t)=f(\varphi(t)), \tag{1}
\end{equation*}
$$

generated by some function $\varphi:[0,1] \rightarrow[0,1]$, the second the composition operator (also called outer superposition operator)

$$
\begin{equation*}
C_{g}(f)(x):=(g \circ f)(x)=g(f(x)) \tag{2}
\end{equation*}
$$

generated by some function $g: \mathbb{R} \rightarrow \mathbb{R}$. These operators may be considered as some kind of "twin brothers": in (1) the inner function is fixed and the outer function $f$ varies over some function space, while in (2) the outer function is fixed and the inner function $f$ varies over some function space. There is one essential difference, however: the operator $S_{\varphi}$ in (1) is linear, while the operator $C_{g}$ in (2) is nonlinear (which makes its study pretty complicated). Thus, in contrast to $S_{\varphi}$ we have to distinguish between boundedness and continuity for $C_{g}$, because a nonlinear operator may be bounded and discontinuous, or continuous and unbounded, as is shown, e.g., by Example 6 below.

In the monograph [6], the nonlinear operator (2) has been studied in these and many other functions spaces ${ }^{1}$; however, many (in part completely unexpected) new results have been published by the nonlinear community since then. For example, the monograph [16] contains many remarkable results on composition operators in Sobolev and related spaces which are of utmost importance in the theory and applications of nonlinear partial differential equations.

The purpose of this survey is to discuss the analytical and topological properties of the operators (1) and (2), to provide a comparison of such properties, and to illustrate them by means of examples and counterexamples. Typical questions we are interested in read as follows:

- Under which condition on $\varphi$ is $S_{\varphi}$ injective in $B V$ ?
- Under which condition on $\varphi$ is $S_{\varphi}$ bounded in Lip?
- Under which condition on $\varphi$ is $S_{\varphi}$ an isometry in Lip?
- Under which condition on $\varphi$ is $S_{\varphi}$ compact in $C$ ?
- Under which condition on $\varphi$ is $S_{\varphi}$ compact in $B V$ ?
- Under which condition on $g$ is $C_{g}$ injective in $C$ ?
- Under which condition on $g$ is $C_{g}$ bounded in Lip?
- Under which condition on $g$ is $C_{g}$ continuous in Lip?
- Under which condition on $g$ is $C_{g}$ Lipschitz continuous in Lip?
- Under which condition on $g$ is $C_{g}$ compact in $B V$ ?

Complete answers to all these and many more questions, in the sense of conditions which are both necessary and sufficient, will be given below ${ }^{2}$.

[^0]This survey paper is organized as follows. We start with characterizing some mapping properties of the operators (1) and (2), like injectivity and surjectivity, in terms of the generating functions $\varphi$ resp. $g$. Afterwards we study some of their analytical properties, like boundedness, continuity, and Lipschitz continuity. One of the most important topological properties of an operator is compactness, which we will consider in the last section. Recall that Lipschitz continuity and compactness are the crucial properties of nonlinear operators in applications of standard fixed point principles, like the Banach fixed point theorem, the Schauder fixed point theorem, and their numerous generalizations.

Loosely speaking, our general question may be posed as follows: does $\varphi$ "feel" the properties of $S_{\varphi}$, and does $g$ "feel" the properties of $C_{g}$ ? The answer to this question is sometimes quite easy, sometimes surprisingly difficult, and sometimes simply unknown.

We point out that most theorems presented below are known; this is not unusual for a survey article. Instead, we put the main emphasis on examples and counterexamples which illustrate that the hypotheses of a result are sharp or, in some cases, how far sufficient conditions are from being necessary.

## 2. Mapping Properties

Although this is not our main focus here, we recall in this section some elementary mapping properties of the operators (1) and (2).

The first question is of course to find conditions on $\varphi$ resp. $g$, both necessary and sufficient, under which the operator (1) resp. (2) maps the above spaces into themselves. In other words, we want to find the largest possible class of "changes of variable" $\varphi$ for which the composition $f \circ \varphi$ remains in some space if we take $f$ from that space. Similarly, we want to find the largest possible class of "perturbations" $g$ for which the composition $g \circ f$ remains in some space if we take $f$ from that space. For the spaces $C$ and Lip this is completely trivial: since both spaces are stable under compositions and contain the identity $f(x)=x$, we have $S_{\varphi}(C) \subseteq C$ iff $\varphi \in C$, $C_{g}(C) \subseteq C$ iff $g \in C(\mathbb{R}), S_{\varphi}($ Lip $) \subseteq$ Lip iff $\varphi \in$ Lip, and $C_{g}($ Lip $) \subseteq$ Lip iff $g \in L i p_{l o c}(\mathbb{R})$.

The analogous problem for $B V$ is much harder. It is clear that the monotonicity of $\varphi$ is sufficient for $S_{\varphi}(B V) \subseteq B V$ but it is obviously not necessary. On the other hand, the condition $\varphi \in B V$ is necessary, since the identity $f(x)=x$ has bounded variation, but it is not sufficient, as the following example shows which is taught in every first year calculus course.

Example 1. The function $\varphi(t)=t^{2} \sin ^{2}(1 / t)$ has bounded variation on $[0,1]$, but the corresponding operator (1) does not map $B V$ into itself, since $f \in B V$ for $f(x)=\sqrt{|x|}$, but $S_{\varphi}(f)=$ $f \circ \varphi \notin B V$.

The problem of characterizing the admissible class of functions $\varphi$ was completely solved in a pioneering paper by Josephy [10], where the author introduced a class of functions which we call pseudomonotone in the paper [5]. Pseudomonotonicity means, loosely speaking, that the number of connected components of $\varphi^{-1}(I)$ is bounded for every interval $I$. It is clear that every monotone function $\varphi$ is pseudomonotone, since $\varphi^{-1}(I)$ is again an interval for each interval $I$, and it is easy to find functions which are pseudomonotone but not monotone. On the other hand, one can show that every pseudomonotone function lies in $B V$, but not vice versa, as Example 1 shows. Josephy's main result reads as follows.

Theorem 1 [10]. (a) The substitution operator $S_{\varphi}$ maps $B V$ into itself iff $\varphi$ is pseudomonotone.
(b) The composition operator $C_{g}$ maps $B V$ into itself iff $g$ is locally Lipschitz.

In this introductory section we will also be interested in elementary mapping properties, like injectivity, surjectivity, or bijectivity of $S_{\varphi}$ and $C_{g}$. A stronger property than injectivity is being an isometry, i.e., preserving norms. Even when dealing with such harmless properties one encounters some unexpected features, and it turns out that these properties heavily depend on the function space we are working in. We summarize with the following Tables 1-3 and, later, Tables 4 and 5.

| $S_{\varphi}(C) \subseteq C$ | $S_{\varphi}$ surjective | $S_{\varphi}$ injective | $\Leftrightarrow$ | $S_{\varphi}$ isometry |
| :---: | :---: | :---: | :---: | :---: |
| $\Uparrow$ | $\Uparrow$ | $\Uparrow$ |  | $\Uparrow$ |
| $\varphi$ continuous | $\varphi$ injective | $\varphi$ surjective |  | $\varphi$ surjective |

Table 1: Properties of $S_{\varphi}: C \rightarrow C$

| $S_{\varphi}($ Lip $) \subseteq$ Lip $^{2}$ | $S_{\varphi}$ surjective | $S_{\varphi}$ injective | $\Leftarrow$ | $S_{\varphi}$ isometry |
| :---: | :---: | :---: | :---: | :---: |
| $\Uparrow$ | $\Downarrow$ | $\Uparrow$ | $\Uparrow$ |  |
| $\varphi$ Lipschitz | $\varphi$ injective | $\varphi$ surjective | $\varphi(t)=t$ |  |

Table 2: Properties of $S_{\varphi}: L i p \rightarrow$ Lip

| $S_{\varphi}(B V) \subseteq B V$ | $S_{\varphi}$ surjective | $S_{\varphi}$ injective | $\Leftarrow$ | $S_{\varphi}$ isometry |
| :---: | :---: | :---: | :---: | :---: |
| $\Uparrow$ | $\Downarrow$ | $\Uparrow$ |  | $\Uparrow$ |
| $\varphi$ pseudomonotone | $\varphi$ injective | $\varphi$ surjective |  | $\varphi$ homeomorphism |

Table 3: Properties of $S_{\varphi}: B V \rightarrow B V$
We do not prove the single assertions, since the (nontrivial) proofs may be found in the recent paper [5]. Instead, we make some comments which illustrate both the expected and the unexpected features.

First of all, Table 1 shows that the situation is most satisfactory in the space $C$, since all conditions are both necessary and sufficient. It is interesting to note the "crossover" in Table 1: injectivity and surjectivity change their roles. The subsequent Tables 2 and 3 show that the situation is slightly different in the space $L i p$ and $B V$ : the surjectivity of $S_{\varphi}$ is only sufficient for the injectivity of $\varphi$ but not necessary (see below for counterexamples).

Remarkably, whenever $S_{\varphi}$ is injective in $C$, we get as a fringe benefit that it is even an isometry, which is of course much more. This is easy to see: if $S_{\varphi}$ is injective, $\varphi$ is surjective, and therefore

$$
\left\|S_{\varphi}(f)\right\|_{C}=\max \{|f(\varphi(s))|: 0 \leq s \leq 1\}=\max \{|f(t)|: 0 \leq t \leq 1\}=\|f\|_{C}
$$

Table 2 and Table 3 show that this is different in Lip and $B V$. For the isometry property of $S_{\varphi}$ we also have necessary and sufficient criteria in terms of $\varphi$ in all spaces, but they are all different in the three tables. To be specific, the function $\varphi(t)=t$ generates an isometric substitution operator in $C$, Lip, and $B V$, the function $\varphi(t)=t^{2}$ only in $C$ and $B V$, but not in Lip, and the function $\varphi(t)=4 t(1-t)$ only in $C$, but neither in Lip nor in $B V$.

In the next two tables we compare what we know about mapping properties of composition operators in the spaces $C, L i p$, and $B V$. Since the properties of $C_{g}$ in $L i p$ and $B V$ are the same, we unify them in one table.


Table 4: Properties of $C_{g}$ in $C$

| $C_{g}(X) \subseteq X$ | $C_{g}$ surjective | $C_{g}$ injective |
| :---: | :---: | :---: |
| $\Uparrow$ | $\Downarrow$ | $\Uparrow$ |
| $g \in$ Lip $_{l o c}$ | $g$ surjective | $g$ injective |

Table 5: Properties of $C_{g}$ in $X \in\{L i p, B V\}$
Note that for the operator (2) there is no "crossover" between injectivity and surjectivity in $C$, as for the operator (1). Our tables show that the nonlinear composition operator behaves in rather the same way in the three spaces $C, L i p$, and $B V$, while the behavior of the linear substitution operator is quite different in these spaces. This is somewhat surprising, because usually a nonlinear operator exhibits more peculiarities than a linear operator. The operator $C_{g}$ has the most interesting analytical properties in the space $B V$, as we will show in the next section.

Before passing to such properties, however, we give some examples which show that, whenever we only have an implication in the Tables $1-5$, these implications cannot be inverted.

Example 2. The function $\varphi:[0,1] \rightarrow[0,1]$ defined by $\varphi(t):=t^{2}$ is injective. On the hand, the corresponding operator $S_{\varphi}$ is not surjective in Lip, because the function $h(x):=x$ is not in its range.

Example 3. Constructing an analogous example for $B V$ is much more complicated. The paper [5] contains a sophisticated example (too technical to be reproduced here) of a bijective pseudomonotone function $\varphi:[0,1] \rightarrow[0,1]$ such that $\varphi^{-1}:[0,1] \rightarrow[0,1]$ is not pseudomonotone, and even not of bounded variation. So the corresponding operator $S_{\varphi}$ maps $B V$ into itself, but it is not surjective, because the function $h(x):=x$ is not in its range.

Example 4. The function $g: \mathbb{R} \rightarrow \mathbb{R}$ defined by $g(u):=\min \{u+2,|u|\}$ is piecewise monotone. Geometrically, the graph of $g$ consists of three linear pieces with corner points at $(-1,1)$ and $(0,0)$. This shows that $g$ is (even globally) Lipschitz continuous on $\mathbb{R}$ with Lipschitz constant 1, so the operator $C_{g}$ maps $C$ into $C$ and Lip into Lip. Moreover, $g$ is certainly surjective. On
the other hand, the corresponding operator $C_{g}$ is not surjective, neither in $C$ nor in Lip, which can be seen as follows.

The function $h(x)=3 x-1$ is a Lipschitz continuous homeomorphism between the intervals $[0,1]$ and $[-1,2]$. If $f \in C$ or $f \in \operatorname{Lip}$ is a function satisfying $C_{g}(f)=h$, then $f$ must be injective. On the other hand, since $h(0)=-1<0$ and $h(1)=2>1$, we have $f(t)=h(t)-2$ for $0 \leq t \leq 2 / 3$, but simultaneously $f(t)=h(t)$ for $1 / 3 \leq t \leq 1$, a contradiction.

Example 5. The function $g: \mathbb{R} \rightarrow \mathbb{R}$ defined by $g(u):=u^{3}$ is a homeomorphism with $g \in$ $\operatorname{Lip}_{\text {loc }}(\mathbb{R})$, but $g^{-1} \notin \operatorname{Lip} p_{\text {loc }}(\mathbb{R})$. Clearly, the corresponding composition operator $C_{g}$ is injective in $B V$. However, $C_{g}$ is not surjective. To see this, observe that the function

$$
h(x):= \begin{cases}1 / n^{3} & \text { for } x=1 / n \\ 0 & \text { otherwise }\end{cases}
$$

belongs to $B V$. The only possible preimage $f$ of $h$ satisfies $f(1 / n)=1 / n$ and $f(t)=0$ otherwise, which does not belong to $B V$, since the harmonic series diverges.

## 3. Analytical Properties

Now we turn to the most important analytical properties of operators which are boundedness and continuity. While for the linear operator (1) this is the same, we point out again that these properties are independent for the nonlinear operator (2). It is not hard to see that, whenever the operator $S_{\varphi}$ maps one of our three spaces into itself, it is automatically bounded.

Continuity of the operators (1) and (2) in the respective norm is much more difficult (and therefore more interesting). Proving that the condition $S_{\varphi}(C) \subseteq C$ (hence $\varphi \in C$ ) implies the continuity of $S_{\varphi}$, and the condition $C_{g}(C) \subseteq C$ (hence $g \in C(\mathbb{R})$ ) implies the continuity of $C_{g}$ in the respective norm is rather straightforward. It is not hard to see that the operator $C_{g}$ is, under the hypothesis $g \in \operatorname{Lip}_{l o c}(\mathbb{R})$, always bounded in Lip, see, e.g., [4]. Remarkably, $C_{g}$ need not be continuous; this is in contrast to the situation in $C$.

Example 6 [7]. On the space Lip, consider the composition operator $C_{g}$ generated by the function $g(u):=\min \{|u|, 1\}$. Then $C_{g}$ is bounded in Lip, but discontinuous at $f(t)=t$, because for the sequence $\left(f_{n}\right)_{n}$ with $f_{n}(t):=t+1 / n$ we have

$$
\left|C_{g}\left(f_{n}\right)(1)-C_{g}(f)(1)-C_{g}\left(f_{n}\right)(1-1 / n)+C_{g}(f)(1-1 / n)\right|=\frac{1}{n}
$$

hence

$$
\left\|C_{g}\left(f_{n}\right)-C_{g}(f)\right\|_{L i p} \geq \frac{1 / n}{1 / n}=1
$$

although $\left\|f_{n}-f\right\|_{\text {Lip }} \rightarrow 0$ as $n \rightarrow \infty$.
Example 6 shows that Lipschitz continuity of $g$ does not guarantee the continuity of $C_{g}$ in the space $C_{g}$, so the right condition on $g$ should be stronger. This problem was solved more than 20 years ago by Goebel and Sachweh who proved the following

Theorem 2 [9]. The composition operator $C_{g}$ maps Lip into itself and is continuous in the norm of Lip iff $g \in C^{1}(\mathbb{R})$.

Example 6 shows that a function $g$ which fails to be differentiable at only few points may generate a discontinuous composition operator in Lip. Continuity of $C_{g}$ in $B V$ is remarkable:

Theorem 3. Whenever the composition operator $C_{g}$ maps $B V$ into itself it is automatically continuous in the norm of $B V$.

Theorem 3 has an interesting history. Finding a condition on $g$, both necessary and sufficient, which guarantees the continuity of $C_{g}$ in $B V$, has been an open problem for many years. The first attempt to prove Theorem 3 has been done in [13], but the proof is 20 pages long and probably not correct. The first complete proof was given by Maćkowiak [11], but it is still very long. A rather brief and very elegant proof was given recently by Reinwand [14].

Let us make a brief detour on Lipschitz continuity of our operators. Of course, this problem is irrelevant for $S_{\varphi}$, since a bounded linear operator is always Lipschitz continuous. For the nonlinear operator $C_{g}$, however, one encounters a surprising degeneracy of $g$ when one requires Lipschitz continuity of $C_{g}$. We summarize with the following

Theorem 4. (a) The operator $C_{g}$ is Lipschitz continuous in the norm of the space $C$ iff $g$ is Lipschitz continuous on $\mathbb{R}$.
(b) The operator $C_{g}$ is Lipschitz continuous in the norm of the space Lip iff $g$ is affine on $\mathbb{R}$, i.e., $g(u)=a u+b$ for some $a, b \in \mathbb{R}$.
(c) The operator $C_{g}$ is Lipschitz continuous in the norm of the space BV iff $g$ is affine on $\mathbb{R}$, i.e., $g(u)=a u+b$ for some $a, b \in \mathbb{R}$.

We make some comments on this theorem. Part (a) was proved in [2] and is what we may expect: Lipschitz continuity of $g$ is reflected in Lipschitz continuity of $C_{g}$, and vice versa. Part (b) was proved in [12] and shows that one has to be very careful with hypotheses on composition operators in Lip: in particular, one may apply the Banach fixed point principle only if the underlying problem is actually linear, and therefore not interesting at all. Part (c) shows that the same degeneracy occurs in $B V$; the proof may be found in [3].

We again summarize our results in the following two tables, the first referring to boundedness, the second to (Lipschitz) continuity.

| $S_{\varphi}$ in $C$ | $S_{\varphi}$ in Lip | $S_{\varphi}$ in $B V$ | $C_{g}$ in $C$ | $C_{g}$ in Lip | $C_{g}$ in $B V$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| bounded | bounded | bounded | bounded | bounded | bounded |
| $\Uparrow$ | $\Uparrow$ | $\Uparrow$ | $\Uparrow$ | $\Uparrow$ | $\Uparrow$ |
| always | always | always | always | always | always |

Table 6: Boundedness of $S_{\varphi}$ and $C_{g}$

| $C_{g}$ in $C$ | $C_{g}$ in Lip | $C_{g}$ in $B V$ | $C_{g}$ in $C$ | $C_{g}$ in Lip | $C_{g}$ in $B V$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| continuous | continuous | continuous | Lip continuous | Lip continuous | Lip continuous |
| $\hat{\Downarrow}$ | $\Uparrow$ | $\Uparrow$ | $\Uparrow$ | $\Uparrow$ | $\Uparrow$ |
| always | $g \in C^{1}(\mathbb{R})$ | always | $g \in \operatorname{Lip}(\mathbb{R})$ | $g$ affine | $g$ affine |

Table 7: Continuity of $C_{g}$
Since boundedness, continuity and Lipschitz continuity is the same for linear operators, we only consider the operator $C_{g}$ in Table 7. Here the label "always" means, of course, under the hypothesis that the operator maps the underlying space into itself.

## 4. Topological Properties

Apart from Lipschitz continuity, compactness is another ingredient which is crucial in fixed point theory. So we spend some time in this last section to analyze what compactness of the operators (1) and (2) means in the three spaces we are interested in.

We begin with the linear operator $S_{\varphi}$. Part (a) of the following theorem has been proved in [5], part (b) in [6].

Theorem 5. (a) The operator $S_{\varphi}$ is compact in the space $C$ iff $\varphi$ is constant on $[0,1]$. The same is true in the space Lip.
(b) The operator $C_{g}$ is compact in the space $C$ iff $g$ is constant on $\mathbb{R}$. The same is true in the space Lip.

A several times before, the situation is different in the space $B V$. The following Example 7 shows that $S_{\varphi}$ may be compact in $B V$ even if $\varphi([0,1])$ contains more than one element. On the other hand, the subsequent Example 8 shows that $S_{\varphi}$ may become noncompact in $B V$ if $\varphi([0,1])$ contains "too many elements".

Example 7. Fix finitely many elements $t_{1}, \ldots, t_{m} \in(0,1]$, and define $\varphi:[0,1] \rightarrow[0,1]$ by $\varphi(t):=$ $\chi_{\left\{t_{1}, \ldots, t_{m}\right\}}(t)$. Since $\varphi$ is pseudomonotone, the operator $S_{\varphi}$ maps $B V$ into itself and is bounded. Moreover, $S_{\varphi}$ is even compact, being an operator of finite rank.

Example 8. Let $I_{n}:=(1 /(n+1), 1 / n]$, and define $\varphi:[0,1] \rightarrow[0,1]$ by $\varphi(t):=1 / n$ if $t \in I_{n}$ and $\varphi(0):=0$. Being monotonically increasing, $\varphi$ is again pseudomonotone, therefore $S_{\varphi}$ maps $B V$ into itself and is bounded. However, $S_{\varphi}$ maps the bounded sequence $\left(\chi_{\{1 / n\}}\right)_{n}$ into the sequence $\left(\chi_{I_{n}}\right)_{n}$ which does not contain a convergent subsequence. Therefore $S_{\varphi}$ cannot be compact in $B V$.

Very often the compactness of a linear operator is proved by approximating the operator by a sequence of finite rank operators. It is interesting that there exist classes of operators for which being compact and having finite rank are actually equivalent. An example are projections, which are also compact if and only if they have finite rank, being the identity on their range.

Remarkably, the substitution operator $S_{\varphi}$ has the same property in $B V$, as is shown by the following Theorem 6. Although the proof may be found in the book [15], we provide a proof which gives some insight into the kind of argument that is typical for the space $B V$ and fully explains the Examples 7 and 8 above.

Theorem 6. For a pseudomonotone function $\varphi$, the following three assertions are equivalent.
(a) The operator $S_{\varphi}: B V \rightarrow B V$ has finite rank.
(b) The operator $S_{\varphi}: B V \rightarrow B V$ is compact.
(c) The set $\varphi([0,1])$ is finite.

Proof. It is clear that (a) implies (b). Suppose that (c) is false. Then we may choose infinitely many different points $s_{n}=\varphi\left(t_{n}\right)$ in $\varphi([0,1])$. The sequence $\left(x_{n}\right)_{n}$ with $x_{n}:=\chi_{\left\{s_{n}\right\}}$ is bounded in $B V$, but its image under $S_{\varphi}$ satisfies

$$
\left\|S_{\varphi}\left(x_{m}\right)-S_{\varphi}\left(x_{n}\right)\right\|_{B V} \geq\left|x_{m}\left(\varphi\left(t_{n}\right)\right)-x_{n}\left(\varphi\left(t_{n}\right)\right)\right|=\left|x_{m}\left(s_{n}\right)-x_{n}\left(s_{n}\right)\right|=1
$$

for $m \neq n$, and therefore cannot contain a convergent subsequence. It follows that $S_{\varphi}$ is not compact, so we have shown that (b) implies (c).

Finally, suppose that $\varphi([0,1])$ is finite, say $\varphi([0,1])=\left\{s_{1}, \ldots, s_{m}\right\}$. Then the sets $A_{j}:=$ $\varphi^{-1}\left(\left\{s_{j}\right\}\right)(j=1, \ldots, m)$ form a disjoint covering of $[0,1]$, and each of these sets has only finitely many connected components, because we supposed $\varphi$ to be pseudomonotone. Now, the functions $x_{j}:=\chi_{A_{j}}=S_{\varphi}\left(\chi_{\left\{s_{j}\right\}}\right)(j=1, \ldots, m)$ belong to $B V$ and form a linearly independent set. Moreover, for any $x \in B V$ we have

$$
S_{\varphi}(x)(t)=\sum_{j=1}^{m} x\left(s_{j}\right) \chi_{A_{j}}(t)=\sum_{j=1}^{m} x\left(s_{j}\right) x_{j}(t) \quad(0 \leq t \leq 1) .
$$

We conclude that span $\left(\left\{x_{1}, \ldots, x_{m}\right\}\right)=R\left(S_{\varphi}\right)$, so $S_{\varphi}$ has a finite dimensional range.
Observe that the last part of the proof shows even more: if $\varphi([0,1])$ is finite, the dimension of the range $R\left(S_{\varphi}\right)$ coincides precisely with the number of elements of $\varphi([0,1])$. So the range of the operator $S_{\varphi}$ induced by the function $\varphi$ from Example 7 has dimension $m+1$, and so the dimension may be arbitrarily prescribed.

We point out again the difference between Theorem 5 and Theorem 6: the operator $S_{\varphi}$ is compact in $B V$ if and only if $\varphi([0,1])$ is finite, and compact in $C$ (or Lip) if and only if $\varphi([0,1])$ is a singleton.

Concerning the compactness of the nonlinear operator $C_{g}$, we have the following Theorem 7; the proof may be found in [6].

Theorem 7. The operator $C_{g}$ is compact in the space $C$ iff $g$ is constant on $\mathbb{R}$. The same is true in the spaces Lip and BV.

Here is what we know about the compactness of the operators $S_{\varphi}$ and $C_{g}$ in the spaces $C, L i p$, and $B V$.

| $S_{\varphi}$ in $C$ | $S_{\varphi}$ in Lip | $S_{\varphi}$ in $B V$ | $C_{g}$ in $C$ | $C_{g}$ in Lip | $C_{g}$ in $B V$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| compact | compact | compact | compact | compact | compact |
| $\Uparrow$ | $\hat{\Downarrow}$ | $\hat{\Downarrow}$ | $\hat{\Downarrow}$ | $\Uparrow$ | $\mathbb{\Downarrow}$ |
| $\varphi$ constant | $\varphi$ constant | $\varphi([0,1])$ finite | $g$ constant | $g$ constant | $g$ constant |

Table 8: Compactness of $S_{\varphi}$ and $C_{g}$
If an operator is not compact it is interesting to measure its "degree of noncompactness". This may be done by calculating the measure of noncompactness of the operator; this leads to Darbo's celebrated fixed point principle [8] which has found numerous important applications. For linear operators $A$ in a normed space $X$, there is another way to measure how far $A$ is from being compact, namely its essential norm

$$
\|A\| \|_{X \rightarrow X}:=\inf \left\{\|A-K\|_{X \rightarrow X}: K: X \rightarrow X \text { compact }\right\} .
$$

Of course, $\|\|A\|\|_{X \rightarrow X}$ is nothing else but the distance of $A$ from the closed two-sided ideal $\mathscr{K}(X)$ of all compact operators in the normed space $\mathscr{L}(X)$ of all bounded linear operators in $X$; in particular, $\left\|\|A\|_{X \rightarrow X}=0\right.$ iff $A$ is compact. Equivalently, $\|\|A\| \|_{X \rightarrow X}$ may be viewed as norm of the class of $A$ in the Calkin algebra $\mathscr{L}(X) / \mathscr{K}(X)$.

Now, in the recent paper [1] it is shown that, whenever $\varphi([0,1])$ contains infinitely many point (and hence $S_{\varphi}$ is not compact in $C$ ), we always have $\left\|\mid S_{\varphi}\right\|\left\|_{C \rightarrow C}=\right\| S_{\varphi} \|_{C \rightarrow C}=1$. This surprising result shows that the norm and essential norm of $S_{\varphi}$ exhibit a certain "bang-bang principle" in the space $C$ : if $\varphi$ is constant the essential norm of $S_{\varphi}$ can assume only the value 0 , and if $\varphi$ is not constant, both the norm and the essential norm of $S_{\varphi}$ can assume only the value 1 ; intermediate values are not possible. In particular, the size of $\left\|\mid S_{\varphi}\right\| \|_{C \rightarrow C}$ does not depend on the size of the range $\varphi([0,1])$ of $\varphi$, as one could expect. For example, the function $\varphi_{c}(t):=c t$ $(0<c \leq 1)$ satisfies $\left\|\mid S_{\varphi_{c}}\right\| \|_{C \rightarrow C}=1$, no matter how close to 0 we choose $c$, i.e., how flat we choose the graph of this function, and so the essential norm of $S_{\varphi_{c}}$ does not "feel" the size of $\operatorname{diam} \varphi_{c}([0,1])=c$ and the slope of the graph.

## REFERENCES

[1] L. Angeloni, J. Appell, T. Domínguez Benavides, S. Reinwand, G. Vinti, Compactness properties of multiplication and substitution operators, J. Oper. Theory 89 (2023) 49-74.
[2] J. Appell, Implicit functions, nonlinear integral equations, and the measure of noncompactness of the superposition operator, J. Math. Anal. Appl. 83 (1981) 251-263.
[3] J. Appell, D. Bugajewska, P. Kasprzak, S. Reinwand, Applications of BV Type Spaces, Lect. Notes Nonlin. Anal. 20, J. Schauder Center, Toruń, 2021.
[4] J. Appell, N. Guanda, N. Merentes, J.-L. Sánchez, Boundedness and continuity properties of nonlinear composition operators: A survey, Commun. Appl. Anal. 15 (2011) 153-182.
[5] J. Appell, B. López, S. Reinwand, S. Schöller, Some remarks on substitution and composition operators, Rend. Ist. Mat. Univ. Trieste 53 (2021), 1-25; Corrigendum: Rend. Ist. Mat. Univ. Trieste 53 (2021), 1-2.
[6] J. Appell, P. P. Zabrejko, Nonlinear Superposition Operators, Cambridge Univ. Press, Cambridge, 2012.
[7] M. Z. Berkolajko, On a nonlinear operator acting in generalized Hölder spaces (Russian), Voron. Gos. Univ. Trudy Sem. Funk. Anal. 12 (1969) 96-104.
[8] G. Darbo, Punti uniti in trasformazioni a codominio non compatto, Rend. Sem. Mat. Univ. Padova 24 (1955) 84-92.
[9] M. Goebel, F. Sachweh, On the autonomous Nemytskij operator in Hölder spaces, Zeitschr. Anal. Anw. 18 (1999) 205-229.
[10] M. Josephy, Composing functions of bounded variation, Proc. Amer. Math. Soc. 83 (1981) 354-356.
[11] P. Maćkowiak, On the continuity of superposition operators in the space of functions of bounded variation, Aequationes Math. 91 (2017) 759-777.
[12] J. Matkowski, J. Miś, On a characterization of Lipschitzian operators of substitution in the space $B V([a, b])$, Math. Nachr. 117 (1984) 155-159.
[13] A. P. Morse, Convergence in variation and related topics, Trans. Amer. Math. Soc. 41 (1937), 48-83.
[14] S. Reinwand, Types of convergence which preserve continuity, Real Anal. Exchange 45 (2020) 173-204.
[15] S. Reinwand, Functions of Bounded Variation: Theory, Methods, Applications, Cuvillier-Verlag, Göttingen 2021.
[16] T. Runst, W. Sickel, Sobolev Spaces of Fractional Order, Nemytskij Operators and Nonlinear Partial Differential Equations, DeGruyter, Berlin 1996.


[^0]:    ${ }^{1}$ In the book [6] the main emphasis is put on non-autonomous operators of the form $C_{g}(f)(x)=g(x, f(x))$ which creates several additional difficulties. In this paper we restict ourselves to the autonomous case (2) for not overburdening the presentation. The linear operator (1) is not considered in [6].
    ${ }^{2}$ If you are curious or impatient, here are the answers to the above list: $\varphi$ surjective; $\varphi$ Lipschitz continuous; $\varphi(t)=t ; \varphi$ constant; $\varphi([0,1])$ finite; $g$ injective; always; $g \in C^{1} ; g$ affine; $g$ constant.

