INSTITUTE OF PHYSICS PUBLISHING INVERSE PROBLEMS

Inverse Problems **20** (2004) 977–991 PII: S0266-5611(04)67402-8

Inverse problems for ODEs using contraction maps and suboptimality of the 'collage method'

H E Kunze[1], **J E Hicken**[2] **and E R Vrscay**[2]

[1] Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, N1G 2W1, Canada
[2] Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada

Received 8 August 2003, in final form 28 January 2004
Published 20 May 2004
Online at stacks.iop.org/IP/20/977
DOI: 10.1088/0266-5611/20/3/019

Abstract

Broad classes of inverse problems in differential and integral equations can be cast in the following framework: the optimal approximation of a target x of a suitable metric space X by the fixed point $\bar{x}$ of a contraction map T on X. The 'collage method' attempts to solve such inverse problems by finding an operator T_c that maps the target x as close as possible to itself. In the case of ODEs, the appropriate contraction maps are integral Picard operators. In practice, the target solutions possibly arise from an interpolation of experimental data points. In this paper, we investigate the suboptimality of the collage method. A simple inequality that provides upper bounds on the improvement over collage coding is presented and some examples are studied. We conclude that, at worst, the collage method provides an excellent starting point for further optimization, in contrast to more traditional searching methods that must first select a good starting point.

1. Introduction

In inverse problems one generally wishes to find a mathematical system, appropriate for the problem at hand, that admits a known function (or measure, etc) as an approximate solution. We have been concerned with a class of inverse problems that may be treated within the framework of Banach's theorem for contraction mappings on complete metric spaces [10, 13]. Given an appropriate metric space X and a *target* $x \in X$, one seeks a contractive operator $T : X \to X$ with fixed point $\bar{x} \in X$ that approximates x to a suitable accuracy.

This viewpoint of inverse problems arose from early works in fractal image coding [8, 9, 15]. Here, (X, d) is a complete metric space of *image functions*, $u : D \to \mathbf{R}$, where $D \subset \mathbf{R}^2$ (or $\mathbf{Z}^2$) represents the *pixel space* or support of u. The action of a contractive fractal operator T on u is to produce a union of N greyscale-modified and spatially contracted copies

of subsets of u. As a result, the fixed point function $\bar{u} = T\bar{u}$ is locally self-similar: $\bar{u}$ is a union of distorted copies of subsets of itself.

The original motivation of fractal coding lay in the fact that only the parameters defining the operator T need to be stored. The approximation $\bar{u}$ to a target image u is then generated by the iteration sequence $u_n = T^n u_0$, where u_0 is any starting image, for example, a blank computer screen, i.e., $u_0 = 0$. Moreover, excellent approximations to the image u could be obtained by using a very small amount of data to define the fractal operator T, resulting in significant *image data compression*.

In fractal image coding, however, the problem of finding a contractive operator that minimizes the approximation error $d(x, \bar{x})$ is extremely complicated because of the nature of fractal operators and, consequently, the self-similar nature of their fixed point functions. In fact, Ruhl and Hartenstein [18] showed that this type of fractal image coding is NP-hard. For this reason, fractal image coding methods have relied on a reformulation of the inverse problem that is based on the so-called *collage theorem* [1], a simple consequence of Banach's theorem. In these methods, one searches for a contractive map T_c that minimizes the so-called *collage distance* $d(x, Tx)$. In the fractal coding literature, this minimization procedure is known as *collage coding*.

Collage coding is a greedy algorithm [18] as it seeks to construct a self-similar approximation $\bar{u}$ to a target u in one pass. It is also necessarily suboptimal since minimization of $d(x, Tx)$ does not imply minimization of the approximation error $d(x, \bar{x})$. Several researchers, for example [7, 20], tried to improve on the fixed point approximations yielded by fractal collage coding but with very little, if any, gain.

We have already shown [13] that various parameter estimation problems for differential equations (see, for example, [12, 17, 19] and, more recently, [16]) can be cast into the above class of inverse problems involving contraction mappings. In a typical parameter estimation problem, one is given a set (or sets) of data points x_i that are assumed to lie on a solution curve (or sets of curves) $x(t)$ of a differential equation of the form $\dot{x} = f(x)$. The problem is to find $f(x)$, which is assumed to have a functional form suitable for the problem being studied, e.g., the polynomial of degree n in x. For a given choice of f and initial values, a solution curve $\bar{x}(t)$ is then obtained numerically by integration and then compared with the target curve $x(t)$. The search for an optimal f is performed numerically in an appropriate space of parameters, e.g., the set of coefficients $c_0, \ldots, c_n$ that define the polynomial vector field $f(x)$. Various search methods have been employed.

From the viewpoint of contraction maps and fixed points, these parameter estimation problems search for a vector field f that defines a Picard operator T with fixed point $\bar{x}(t)$ that, in turn, approximates the target function $x(t)$ as well as possible. We emphasize that most of the methods in the literature concentrate on minimizing the approximation error $d(x - \bar{x})$ while the collage method instead minimizes the collage distance $d(x, Tx)$, a useful change since one cannot find $\bar{x}$ for a general T.

In [13], we showed that collage coding is naturally implemented into an inverse problem scheme for ODEs based on the Picard contraction mapping. As well, we showed rigorously that the collage method can be performed in $\mathcal{L}^2$, which facilitates numerical computations. After writing that paper, however, we discovered that a collage coding method had, in fact, been used for simple ODE problems, e.g. [11], although justification of this method in terms of contraction maps as well as the use of the $\mathcal{L}^2$ norm was not acknowledged until [13].

The purpose of this paper is to investigate the suboptimality of collage coding for inverse problems involving ODEs as well as to consider some additional applications that were not discussed in [13]. We first present a very simple inequality that provides upper bounds on the improvement over collage coding, whether it be fractal or non-fractal. Indeed,

this inequality explains the low gains over collage coding obtained in earlier fractal image coding investigations. We then examine a few simple ODE inverse problems and observe that the collage method, although theoretically suboptimal, nevertheless yields excellent approximations.

The structure of this paper is as follows. In section 2, we review the basic mathematical results that lie behind the inverse approximation of fixed points of contraction maps. The application of this method to systems of first-order ODEs is then discussed briefly. We then examine an interesting inverse problem for damped harmonic oscillations as originally formulated by Groetsch [11]—in terms of a second-order integral operator—and show, as expected, that Groetsch's method is precisely collage coding. In section 3, we examine the suboptimality of collage coding and present our simple inequality. We then apply this result to some inverse problems for ODEs, including Groetsch's example as well as an inverse problem in ecological modelling that has been studied in the literature [16]. Essentially the conclusion of this paper is that collage coding for ODE inverse problems performs very well, as is the case for fractal coding. More importantly—in contrast to the usual search methods performed in the literature—it provides an excellent starting point from where additional searching could be performed. In practice, however, the improvements are generally so small that it is generally not worth the effort to even try to compute them.

2. Some mathematical background

In this section, we restrict our discussion of technical details to a minimum. Let (X, d) denote a complete metric space and $\mathrm{Con}(X)$ an appropriate set of contraction maps on X: if $T \in \mathrm{Con}(x)$ then $T : X \to X$ and there exists a $c \in [0, 1)$ such that

$$d(Tx, Ty) \leqslant cd(x, y) \qquad \forall x, y \in X. \tag{1}$$

(In this discussion, we are not concerned about the properties of the set $\mathrm{Con}(X)$.) From Banach's fixed point theorem, there is a unique $\bar{x} \in X$ such that $T\bar{x} = \bar{x}$.

Let $FP(X)$ denote the set of all fixed points of the contraction maps in $\mathrm{Con}(X)$, i.e.

$$FP(X) = \{\bar{x} \in X \mid \bar{x} = T\bar{x} \text{ for some } T \in \mathrm{Con}(X)\}.$$

We now seek to approximate a target element $x \in X$ by fixed points $\bar{x} \in FP(X)$. The error of this approximation will be given by $d(x, \bar{x})$. The optimal fixed point, $\bar{x}_o$ (assuming that at least one exists), minimizes the error, i.e.

$$d(\bar{x}_o, x) \leqslant d(\bar{x}, x) \qquad \forall \bar{x} \in FP(X). \tag{2}$$

As mentioned in the introduction, the direct determination of optimal fixed points is generally infeasible. Instead, one usually resorts to collage coding. The basis of this method rests in the *collage theorem* [1], a simple consequence of Banach's theorem.

Proposition 1. *Let $x \in X$ and $T \in \mathrm{Con}(X)$ with fixed point $\bar{x}$ and contraction factor $c \in [0, 1)$. Then*

$$d(x, \bar{x}) \leqslant \frac{1}{1-c} d(x, Tx). \tag{3}$$

Proof.

$$\begin{aligned} d(x, \bar{x}) &\leqslant d(x, Tx) + d(Tx, \bar{x}) \\ &\leqslant d(x, Tx) + cd(x, \bar{x}). \end{aligned}$$

A rearrangement yields the desired result. □

Collage coding seeks to minimize the *collage distance* $d(x, Tx)$: by making the collage distance small, the approximation error $d(x, \bar{x})$ is made small (subject to some controls that we may have to put on the contraction factors c, for example, $c \leqslant 0.9$).

There is an interesting *anti-collage theorem* [20] that yields a lower bound to the approximation error in terms of the collage error. Once again, it is a simple consequence of Banach's theorem:

Proposition 2. *Let $x \in X$ and $T \in \mathrm{Con}(X)$ with fixed point $\bar{x}$ and contraction factor $c \in [0, 1)$. Then*

$$d(x, \bar{x}) \geqslant \frac{1}{1+c} d(x, Tx). \tag{4}$$

Proof.

$$\begin{aligned} d(x, Tx) &\leqslant d(x, \bar{x}) + d(\bar{x}, Tx) \\ &\leqslant d(x, \bar{x}) + cd(x, \bar{x}). \end{aligned}$$

A rearrangement yields the desired result. □

To summarize, we have the following important bounds on the collage distance:

$$\frac{1}{1+c} d(x, Tx) \leqslant d(x, \bar{x}) \leqslant \frac{1}{1-c} d(x, Tx). \tag{5}$$

2.1. The inverse problem for systems of first-order ODEs

We seek an ODE initial value problem,

$$\dot{x}(t) = f(x, t), \qquad x(0) = x_0, \tag{6}$$

that admits a target solution $x(t)$ as an exact or approximate solution, where f is restricted to a class of functional forms, e.g. affine, quadratic. Associated with the initial value problem in equation (6) is the Picard integral operator T:

$$(Tu)(t) = x_0 + \int_0^t f(u(s), s)\, \mathrm{d}s. \tag{7}$$

It is well known [5] that, subject to appropriate conditions on the vector field f, the operator T is contractive over an appropriate Banach space $\mathcal{C}(I)$ of functions supported over an interval I containing the point $t_0 = 0$. As well, the fixed point $\bar{u}$ of T is the unique solution to the IVP in (6). In [13], we showed that T is also contractive in the $\mathcal{L}^2$ metric, which is much more convenient to work with. Given a target function $x(t), t \in [0, 1]$, say, the $\mathcal{L}^2$ collage distance has the form

$$\Delta = \left(\int_0^1 (x(t) - (Tx)(t))^2 \, \mathrm{d}t \right)^{\frac{1}{2}} \tag{8}$$

$$= \left(\int_0^1 \left[x(t) - x_0 - \int_0^t f(x(s), s)\, \mathrm{d}s \right]^2 \mathrm{d}t \right)^{\frac{1}{2}}. \tag{9}$$

Considering $f(x(s), s)$ to be a function of certain parameters, one can then minimize the collage distance Δ in equation (9). For example, if we assume that f is an autonomous polynomial vector field,

$$f(x) = \sum_{k=0}^{N} c_k x^k, \tag{10}$$

then the minimization of the squared $\mathcal{L}^2$ collage distance yields a set of simultaneous linear equations in the unknowns c_k. (We may also consider the initial value x_0 as an unknown, in which case an additional linear equation appears.) The coefficients of this linear system involve generalized moments of the target function $x(t)$ having the form

$$g_k(t) = \int_0^t [x(s)]^k \, \mathrm{d}s. \tag{11}$$

The parameter values solving this system define a Picard operator T_c with attractor function $\bar{x}_c(t)$. We shall refer to $\bar{x}_c(t)$ as the *collage attractor*. For details, we refer the reader to [13].

Finally, we mention that the minimization of the collage distance along with the condition that f belong to some restricted class of functions represents a kind of regularization of the inverse problem associated with equation (6).

2.2. An inverse problem in structural dynamics

In section 3.4 of [11], the following parameter estimation problem for a damped harmonic oscillator is considered. From some observed measurements $(t_i, x(t_i))$ and knowledge of the initial position $x(0) = x_0$ and velocity $\dot{x}(0) = v_0$, one wishes to determine the coefficients a and b of the second-order linear ODE

$$\ddot{x} + a\dot{x} + bx = 0, \qquad x(0) = x_0, \qquad \dot{x}(0) = v_0, \tag{12}$$

which admits a solution $x(t)$ that interpolates the above data points as closely as possible. This problem can easily be recast into the form of example 1 by transforming the second-order DE into a system of first-order DEs. However, in [11], equation (12) is integrated twice. The resulting integral equation along with the sample measurements leads to a least-squares minimization problem to determine a and b. We now show that this method is precisely collage coding for a contractive operator.

Let us consider the following family of second-order ODEs that includes equation (12) as a special case:

$$\ddot{x}(t) = f(t, x(t), \dot{x}(t)), \qquad x(0) = x_0, \qquad \dot{x}(0) = v_0. \tag{13}$$

Integrating twice gives

$$x(t) = x_0 + v_0 t + \int_0^t (t - s) f(s, x(s), \dot{x}(s)) \, \mathrm{d}s, \qquad t > 0. \tag{14}$$

In other words, the solution $x(t)$ is the fixed point of the integral Picard operator T defined by

$$(Tu)(t) = x_0 + v_0 t + \int_0^t (t - s) f(s, u(s), \dot{u}(s)) \, \mathrm{d}s. \tag{15}$$

Now define

$$I = [0, \delta], \qquad \delta > 0,$$

$$d_\infty(x_1, x_2) = \sup_{t \in I} \{|x_1 - x_2| + |\dot{x}_1(t) - \dot{x}_2(t)|\}, \qquad \forall x_1, x_2 \in \bar{C}^1(I),$$

$$d_2(x_1, x_2) = \left(\int_0^\delta (|x_1 - x_2| + |\dot{x}_1(t) - \dot{x}_2(t)|)^2 \, \mathrm{d}t \right)^{\frac{1}{2}}, \qquad \forall x_1, x_2 \in \bar{C}^1(I),$$

$$\bar{C}^1(I) = \{x \in C^1(I) \mid \|x\|_\infty \leqslant M\} \qquad \text{and} \qquad D = \{(t, x, \dot{x}) \mid t \in I, \|x\|_\infty \leqslant M\}.$$

In appendix A, we prove the following result.

Theorem 3. *Suppose that f satisfies*

(i) $\max_{(t,x,\dot{x})\in D} |f(t,x,\dot{x}(t))| \leqslant \frac{M}{(\delta+1)\delta}$ *and*

(ii) the following Lipschitz condition on D: for all $(t_1, x_1, \dot{x}_1)$ *and* $(t_2, x_2, \dot{x}_2)$ *in D, there exist non-negative real numbers* K_1 *and* K_2*, not both zero, such that* $|f(t_1, x_1, \dot{x}_1) - f(t_2, x_2, \dot{x}_2)| \leqslant K_1|x_1 - x_2| + K_2|\dot{x}_1 - \dot{x}_2|$.

Let $K = \max(K_1, K_2) > 0$.

(a) Define

$$\|x\|_{\infty,\lambda} = \sup_{t\in I} \mathrm{e}^{-\lambda K t}|x(t)|.$$

Then the Picard operator T in (15) is contractive on the Banach space $(\bar{C}^1(I), \|\cdot\|_{\infty,2(\delta+1)})$.

(b) Define

$$\|x\|_{2,\lambda} = \left(\int_I (\mathrm{e}^{\lambda K t}x(t))^2\,\mathrm{d}t\right)^{\frac{1}{2}}.$$

Then the Picard operator T in (15) is contractive on $(\bar{C}^1(I), \|\cdot\|_{2,(\delta+1)^2\delta K})$.

The norms in the theorem are often referred to as 'Bielecki norms' due to Bielecki [3]. Given the conditions on f in the theorem, the weighting allows one to establish an existence result regardless of δ, using the fact that such a weighted norm is equivalent to its associated standard norm:

$$\|x\|_{\cdot,\lambda} \leqslant \|x\|_{\cdot} \leqslant \mathrm{e}^{\lambda K\delta}\|x\|_{\cdot,\lambda}, \tag{16}$$

However, using the weighted $\mathcal{L}^2$ norm in the collage theorem, for example, will in general lead to a quite complicated minimization problem for parameters of the differential equation because the Lipschitz constant K will depend upon the parameters of f. The most convenient norm to work with is the standard $\mathcal{L}^2$ norm, even though contractivity in this norm depends on δ. From equation (16), we see that if the $\mathcal{L}^2$ collage distance $d_2(x, Tx) < \varepsilon$, then the associated weighted collage distance $d_{2,\lambda}(x, Tx) < \varepsilon$.

We now return to the damped linear oscillator problem of equation (12). In [11], the n observed data points (t_i, x_i) are assumed to be evenly spaced in time, with $t_i = ih$, $1 \leqslant i \leqslant n$. The integral in the second-order Picard operator is then integrated by parts and the resulting integrals are approximated with the trapezoid rule. The squared $\mathcal{L}^2$ collage distance $\|x - Tx\|_2^2$ is given by the expression (using the notation of [11])

$$E(a,b) = \sum_{k=1}^{n}(E_k(a,b))^2, \tag{17}$$

where

$$E_k(a,b) = x_k - x_0 - v_0kh + a\left(\sum_{j=1}^{k} x_jh - \frac{x_kh}{2} - x_0kh + \frac{x_0h}{2}\right) + b\left(\sum_{j=1}^{k-1} x_j(k-j)h^2 + \frac{x_0kh^2}{2}\right). \tag{18}$$

(Note that there is a term missing in the expression for E_k on p 58 in [11].) Minimization of $E(a,b)$ yields a set of linear equations in a and b.

As Groetsch qualifies in [11], the use of the trapezoidal rule in the Picard integral operator is a very simple approximation. Nevertheless, it yields rather good results. We have investigated the use of other interpolation schemes. Not surprisingly, better estimates

of the parameters are yielded with more sophisticated interpolation schemes. We also finally mention that, if necessary, the initial velocity v_0 could be considered as an unknown parameter. The extension of the above method to determine an optimal v_0 value along with a and b is straightforward.

3. Suboptimality of collage coding

In general, suppose that we perform collage coding of the target $x \in X$, i.e., we find a mapping $T_c \in \mathrm{Con}(X)$ (assuming that at least one exists) that minimizes the collage distance $d(x, Tx)$. Then

$$d(x, T_c x) \leqslant d(x, Tx) \qquad \forall\, T \in \mathrm{Con}(X).$$

Let $\bar{x}_c \in FP(X)$ denote the fixed point of T_c, i.e. $T_c\bar{x}_c = \bar{x}_c$. We shall refer to $\bar{x}_c$ as the *collage attractor*. By our definition of $\bar{x}_o$ in equation (2),

$$d(x, \bar{x}_o) \leqslant d(x, \bar{x}_c). \tag{19}$$

In other words, collage coding is suboptimal.

The following result establishes an upper bound to the distance between $\bar{x}_o$ and $\bar{x}_c$.

Proposition 4.

$$d(\bar{x}_o, \bar{x}_c) \leqslant \frac{2}{1 - c_c} d(x, T_c x), \tag{20}$$

where c_c is the contraction factor of T_c.

Proof.

$$\begin{aligned} d(\bar{x}_o, \bar{x}_c) &\leqslant d(\bar{x}_o, x) + d(x, \bar{x}_c) \\ &\leqslant d(\bar{x}_c, x) + d(x, \bar{x}_c). \end{aligned}$$

The desired result follows from the collage theorem, cf equation (3). □

Note that $\bar{x}_o$ may be replaced by $\bar{x}'$, any fixed point that approximates x no worse than $\bar{x}_c$ does, i.e., $d(x, \bar{x}') \leqslant d(x, \bar{x}_c)$. Then

$$d(\bar{x}', \bar{x}_c) \leqslant \frac{2}{1 - c_c} d(x, T_c x). \tag{21}$$

In other words, once a collage attractor $\bar{x}_c$ corresponding to a contraction map T_c is found, all fixed points $\bar{x} \in FP(X)$ that approximate x as well as $\bar{x}_c$ lie in a closed ball of radius $\frac{2}{1-c_c} d(x, T_c x)$ centred at $\bar{x}_c$. Note that we can make this radius arbitrarily large by allowing the contraction factors c of the maps in $\mathrm{Con}(X)$ to approach 1.

The improvement in the optimal attractor error from the collage error can also be bounded as follows:

Proposition 5.

$$0 \leqslant d(x, \bar{x}_c) - d(x, \bar{x}_o) \leqslant \frac{2}{1 - c_c} d(x, T_c x), \tag{22}$$

where c_c is the contraction factor of T_c.

The proof of this result follows immediately from proposition 4.

In the fractal image coding literature there have been some attempts to improve the results of collage coding by starting with the collage attractor $\bar{x}_c$ and searching for attractors $\bar{x} \in FP(X)$ that lower the approximation error $d(x, \bar{x})$ [7, 20]. These searches are performed

Table 1. Minimal collage and near-optimal parameters for $x(t) = (t+1)^2$, linear f.

	f	x_0	$\|x - Tx\|_2$	$\|x - \bar{x}\|_2$
x_0 constrained, collage	1.403 641 88 + 0.690 440 06x	1	0.005 806 99	0.006 057 84
x_0 constrained, near-optimal	1.403 056 56 + 0.689 931 49x	1	0.005 924 01	0.005 878 51
x_0 variable, collage	1.471 518 99 + 0.664 556 96x	0.988 396 62	0.004 192 96	0.004 183 00
x_0 variable, near-optimal	1.471 472 53 + 0.664 547 04x	0.988 432 82	0.004 193 00	0.004 182 67

in the *fractal code space* $\Pi \subset \mathbf{R}^n$ of parameters that define the fractal operator T (subject to the condition that T be contractive). Implicitly assumed in these studies is the continuity of the respective attractor functions $\bar{x}$ with respect to these parameters, originally proved in [4].

For example, Dudbridge and Fisher [7] used a Nelder–Mead simplex algorithm to perform such a search. Some improvements in the approximation error $d(x, \bar{x})$ were found but they were generally quite small. In addition, the fixed point approximations yielded by this search were not very far away from the collage attractor $\bar{x}_c$ (as expected by the continuity property of the attractors). In [20], gradient descent methods were used to perform the search. However, no significant improvements were found over the simplex algorithm search of [7]. (In [20], the differentiability of attractor functions $\bar{u}$ with respect to fractal parameters was first established in order to justify the use of gradient descent methods. This, in itself, was a very interesting mathematical result, especially since the partial derivatives are *vector functions*. Nevertheless, there was no practical advantage to using gradients.)

3.1. Application to inverse problems for ODEs

Without loss of generality, we focus our discussion on the inverse problem of section 2.1, using polynomial vector fields as defined in equation (10). The Picard integral operators T and their corresponding attractor functions $\bar{x}(t)$ are implicitly defined by the parameters c_k of equation (10). In other words, the c_k (subject to conditions that ensure the contractivity of the T—we skip the complicated details here (see [13])) form the parameter space Π over which any optimization would have to be performed. As mentioned earlier, minimization of the squared $\mathcal{L}^2$ collage distance in equation (9) yields the collage attractor $\bar{x}_c(t)$. We now seek to find attractor functions $\bar{x}(t)$ that approximate a target function $x(t)$ better than $\bar{x}_c(t)$ does.

Unless one works with a simple choice of the vector field $f(x)$, so that the general solution of the ODE(s) in terms of the parameters in $f(x)$ is readily available, a gradient approach to this problem would be computationally intensive. In general, the computation of gradients of Picard integral operators with respect to the c_k is quite cumbersome, a situation similar to that of fractal coding [20]. (The computation of these gradients is discussed in appendix B.) And at each step in the iterative gradient descent process, full knowledge of the current approximation/solution is required. (Practically speaking, such a solution would be typically known numerically as a discrete sequence.) Therefore, in the study presented below we have employed a simple Nelder–Mead-type parameter search to seek out local attractors that yield better approximations to a target x.

Let us consider the inverse ODE problem of section 2.1, with target function $x(t) = (t+1)^2$ on $[0, 1]$. Note that this target satisfies the initial value problem $\dot{x}(t) = 2\sqrt{x}$, $x(0) = 1$. We first look for a linear initial value problem of the form

$$\dot{x} = c_0 + c_1 x, \qquad x(0) = x_0,$$

with solution as close as possible to our target function. The results obtained by minimizing the $\mathcal{L}^2$ collage distance $\|x - Tx\|_2$ are presented in table 1. There are two cases: (i) treating

$x_0 = 1$ as fixed, (ii) treating x_0 as a variable to be optimized. (The entries are presented to eight decimal places to permit an easy comparison.) Table 1 also presents the near-optimal results obtained by performing, in each case, a simplex parameter search starting at the respective collage attractors $x_c(t)$. In each case, the approximation error is decreased although, as expected, the collage distance increases. In all cases, the improvement over collage coding is rather small.

In the case of linear vector fields, one can actually attempt to solve the direct problem by finding the best $\mathcal{L}^2$ approximation

$$(t+1)^2 \cong A + B\exp(Ct), \qquad t \in [0,1]. \tag{23}$$

The values A, B, C determine the parameters c_0, c_1 and x_0. However, the optimal parameters A, B, C cannot be determined in closed algebraic form, so we must resort to numerical methods.

In the case when x_0 is treated as a parameter, we find, to eight decimal places, that

$$A = -2.214\,248\,86, \qquad B = 3.202\,681\,68, \qquad C = 0.664\,547\,04. \tag{24}$$

These values coincide with the near-optimal result of table 1 (fourth row), where

$$A = -\frac{c_0}{c_1}, \qquad B = x_0 + \frac{c_0}{c_1}, \qquad C = c_1. \tag{25}$$

The upper bound appearing in propositions 4 and 5 may now be computed:

$$\frac{2}{1-c_c}\|x - T_c x\|_2 = \frac{2(0.004\,192\,96)}{1 - \frac{1}{2}\frac{105}{158}} = 0.012\,559\,01, \tag{26}$$

where

$$c_c = c_1 = \frac{105}{158} = 0.664\,556\,96.$$

We expect this bound not to be sharp. Indeed, we observe that the improvement to collage coding,

$$d(x, \bar{x}_c) - d(x, \bar{x}_o) = 0.000\,000\,33,$$

is smaller than this bound by more than four orders of magnitude! We expect a similar lack of sharpness for the approximation error $d(\bar{x}_o, \bar{x}_c)$.

Each of the methods described above yields a vector field $f(x)$ for which the target function $x(t)$ is an approximate solution to the DE $\dot{x} = f(x)$. In the examples considered above, the vector fields are good approximations to the true vector field. This is not surprising since the original proof of the collage method [13] relied on the Weierstrass approximation of $f(x)$. To illustrate, the target function $x(t) = (t+1)^2$ satisfies the IVP,

$$\dot{x} = 2\sqrt{x}, \qquad x(0) = 1. \tag{27}$$

The best $\mathcal{L}^2$ linear approximation to the vector field $2\sqrt{x}$ over the interval $1 \leqslant x \leqslant 4$ is given by

$$g(x) = \frac{40}{27} + \frac{88}{135}x \cong 1.481 + 0.652x. \tag{28}$$

The vector fields listed in table 1 are seen to be quite close to this result.

We now consider the case of quadratic vector fields for this problem, i.e.,

$$\dot{x} = c_0 + c_1 x + c_2 x^2, \qquad x(0) = x_0, \tag{29}$$

i.e., $N = 2$ in equation (10). The results are presented to nine decimal places in table 2 in order to permit an easy comparison. Once again, it is seen that the improvements to collage coding are very small.

Table 2. Collage and near-optimal parameters for $x(t) = (t+1)^2$, quadratic f.

	f	x_0	$\|x - Tx\|_2$	$\|x - \bar{x}\|_2$
x_0 constrained, collage	$1.060\,837\,408 + 1.038\,471\,231x - 0.078\,366\,472x^2$	1	0.000 695 750	0.000 719 857
x_0 constrained, near-optimal	$1.060\,617\,408 + 1.038\,269\,231x - 0.078\,265\,472x^2$	1	0.000 704 726	0.000 702 958
x_0 variable, collage	$1.092\,978\,223 + 1.012\,130\,627x - 0.073\,287\,330x^2$	0.988 303 740	0.000 527 977	0.000 526 460
x_0 variable, near-optimal	$1.092\,965\,623 + 1.012\,121\,103x - 0.073\,281\,930x^2$	0.988 308 840	0.000 527 981	0.000 526 454

The best $\mathcal{L}^2$ quadratic approximation to the vector field $2\sqrt{x}$ over the interval $1 \leqslant x \leqslant 4$ is given by

$$g(x) = \frac{620}{567} + \frac{2848}{2835}x - \frac{40}{567}x^2 \cong 1.093 + 1.005x - 0.071x^2. \tag{30}$$

The vector fields listed in table 1 are seen to be quite close to this result.

Finally, we examine the inverse problems studied in [16], wherein the parameter values of certain ecological models are estimated with a Nelder–Mead-type search method. In that paper, 'synthetic data' are generated by numerically solving the differential equations of a proposed model with specified parameters. Gaussian noise is added to the numerical solution, which is then sampled at a number of uniformly distributed points. These sample points are fed into the parameter estimation process outlined in [16] in order to determine optimal parameter values for differential equations of the proposed form. We emphasize that, to the best of our understanding, the estimation method in [16] involves the minimization of the fixed point approximation error $\|x - \bar{x}\|$, as opposed to minimization of the collage error $\|x - Tx\|$.

One particular case studied in [16] is the SML model,

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -K_s SX, \tag{31}$$

$$\frac{\mathrm{d}X}{\mathrm{d}t} = K_\mathrm{c} SX - K_m \frac{X^2}{S}, \tag{32}$$

where $S(t)$ and $X(t)$ represent the substrate concentration and biomass at time t and the parameters K_s, K_c and K_m are all positive. Of course, the Nelder–Mead-type search of [16] determines the near-optimal parameters for system (31)–(32) to have a solution as close as possible to the noised numerical solution. Increasing the standard deviation of the noise distribution decreases the quality of the fit to the parameters for the system with zero-noise solution. The Gaussian distribution had zero mean and peak magnitude as large as the peak value of the variable to which it was added.

We have employed collage coding on this problem using the same test parameters employed in [16]: $K_b = 0.0055$, $K_\mathrm{c} = 0.0038$ and $K_m = 0.000\,55$. The system was solved numerically: 100 sampled data points (with Gaussian noise of low-amplitude ε added) were fitted to a 10-degree polynomial and collage coding was then employed. The sampled data points for $S(t)$ were also used to fit the function $S(t)^{-1}$ that appears in (32).

We make two comments before presenting the results in table 3: (i) Nelder–Mead-type searches are typically quite time- and resource-consuming, while the collage coding method

Table 3. Collage coding results for the SML problem of [16].

ε	K_b	K_c	K_m
0.00	0.005 500 000	0.003 800 000	0.000 550 000
0.01	0.005 520 150	0.003 739 329	0.000 531 034
0.03	0.005 560 584	0.003 617 699	0.000 492 936
0.05	0.005 601 198	0.003 495 669	0.000 454 616
0.10	0.005 703 507	0.003 188 687	0.000 357 837

is quite fast, and (ii) nowhere in [16] is the initial guess at the parameters, i.e. the seed for the algorithm, specified. This second point is quite important, as the results of the collage coding method provide an excellent initial guess for further optimization, which may not even be warranted because the results are so close.

4. Concluding remarks

In this paper we have investigated the suboptimality of collage coding for inverse problems for ODEs. We first presented a very simple inequality that provides an upper bound on the improvement over collage coding. Indeed, this inequality explains the low improvements to collage coding obtained in earlier investigations of fractal image coding. In the case of ODEs, the improvements to collage coding are found to be very small. We conjecture that any negative effects of the 'greediness' of the collage coding method for ODEs are minimal since solutions to ODEs—also fixed points to contractive Picard integral operators—possess a great deal of regularity. This is in contrast to the complexity of self-similar functions that are fixed points of fractal operators. An important consequence is that the collage method provides excellent approximations in one procedure, as opposed to more traditional methods that must first select a good starting point before undertaking a searching procedure.

Acknowledgments

We gratefully acknowledge the support of this research by the Natural Sciences and Engineering Research Council of Canada (NSERC) in the form of individual Grants in Aid of Research (HEK and ERV) and an NSERC Undergraduate Student Research Assistantship (JEH).

Appendix A

In this appendix, we prove theorem 3, establishing the contractivity of the Picard operator (15) in both the weighted sup and $\mathcal{L}^2$ norms.

Without loss of generality, setting $x_0 = \dot{x}_0 = 0$, we have

$$\begin{aligned}
\|Tx\|_\infty &= \sup_{t\in I}\left(\left|\int_0^t (t-s)\,f(s,x(s),\dot{x}(s))\,\mathrm{d}s\right| + \left|\int_0^t f(s,x(s),\dot{x}(s))\,\mathrm{d}s\right|\right)\\
&\leqslant \sup_{t\in I}\int_0^t (|t-s|+1)|f(s,x(s),\dot{x}(s))|\,\mathrm{d}s\\
&\leqslant \frac{M}{(\delta+1)\delta}\sup_{t\in I}\int_0^t (\delta+1)\,\mathrm{d}s\\
&\leqslant M
\end{aligned}$$

and

$$\frac{\mathrm{d}(Tx)}{\mathrm{d}t}(t) = \int_0^t f(s, x(s), x'(s))\,\mathrm{d}s \in C(I).$$

By this construction, $T : \bar{C}^1(I) \mapsto \bar{C}^1(I)$.

We first prove theorem 3(a), contractivity with respect to the weight sup norm.

Proof. For $t \in I$,

$$\begin{aligned}
|Tx_1 - Tx_2| + \left|\frac{\mathrm{d}(Tx_1)}{\mathrm{d}t} - \frac{\mathrm{d}(Tx_2)}{\mathrm{d}t}\right| &= \left|\int_0^t (t-s)\left(f\left(s, x_1(s), \frac{\mathrm{d}x_1}{\mathrm{d}s}(s)\right) - f\left(s, x_2(s), \frac{\mathrm{d}x_2}{\mathrm{d}s}(s)\right)\right)\mathrm{d}s\right| \\
&\quad + \left|\int_0^t f\left(s, x_1(s), \frac{\mathrm{d}x_1}{\mathrm{d}s}(s)\right) - f\left(s, x_2(s), \frac{\mathrm{d}x_2}{\mathrm{d}s}(s)\right)\mathrm{d}s\right| \\
&\leqslant \int_0^t (|t-s|+1)\left|f\left(s, x_1(s), \frac{\mathrm{d}x_1}{\mathrm{d}s}(s)\right) - f\left(s, x_2(s), \frac{\mathrm{d}x_2}{\mathrm{d}s}(s)\right)\right|\mathrm{d}s.
\end{aligned}$$

Now use the fact that $0 \leqslant |t-s| \leqslant t \leqslant \delta$ and apply the Lipschitz condition (ii) of the theorem.

$$\begin{aligned}
|(Tx_1) - (Tx_2)| + \left|\frac{\mathrm{d}(Tx_1)}{\mathrm{d}t} - \frac{\mathrm{d}(Tx_2)}{\mathrm{d}t}\right| &\leqslant \int_0^t (\delta+1)\left|f\left(s, x_1(s), \frac{\mathrm{d}x_1}{\mathrm{d}s}(s)\right) - f\left(s, x_2(s), \frac{\mathrm{d}x_2}{\mathrm{d}s}(s)\right)\right|\mathrm{d}s \\
&\leqslant (\delta+1)\int_0^t \left(K_1|x_1(s) - x_2(s)| + K_2\left|\frac{\mathrm{d}x_1}{\mathrm{d}s}(s) - \frac{\mathrm{d}x_2}{\mathrm{d}s}(s)\right|\right)\mathrm{d}s \\
&\leqslant (\delta+1)K\int_0^t \mathrm{e}^{\lambda K s}\,\mathrm{e}^{-\lambda K s}\left(|x_1(s) - x_2(s)| + \left|\frac{\mathrm{d}x_1}{\mathrm{d}s}(s) - \frac{\mathrm{d}x_2}{\mathrm{d}s}(s)\right|\right)\mathrm{d}s \\
&\leqslant (\delta+1)K d_{\infty,\lambda}(x_1, x_2)\int_0^t \mathrm{e}^{\lambda K s}\,\mathrm{d}s \\
&\leqslant (\delta+1)K d_{\infty,\lambda}(x_1, x_2)\frac{1}{\lambda K}(\mathrm{e}^{\lambda K t} - 1) \\
&\leqslant \frac{\delta+1}{\lambda}\,\mathrm{e}^{\lambda K t} d_{\infty,\lambda}(x_1, x_2).
\end{aligned}$$

Hence, for $t \in I$,

$$\mathrm{e}^{-\lambda K t}\left(|(Tx_1) - (Tx_2)| + \left|\frac{\mathrm{d}(Tx_1)}{\mathrm{d}t} - \frac{\mathrm{d}(Tx_2)}{\mathrm{d}t}\right|\right) \leqslant \frac{\delta+1}{\lambda} d_{\infty,\lambda}(x_1, x_2)$$

Setting $\lambda = 2(\delta+1)$, we conclude that

$$\|Tx_1 - Tx_2\|_{\infty,2(\delta+1)} \leqslant \tfrac{1}{2}\|x_1 - x_2\|_{\infty,2(\delta+1)}.$$

□

We first prove theorem 3(a), contractivity with respect to the weight sup norm.

Proof. Borrowing from the previous proof, we have

$$\begin{aligned}
\|(Tx_1) - (Tx_2)\|_{2,\lambda}^2 &= \int_I \mathrm{e}^{-2\lambda K t}\left[|Tx_1 - Tx_2| + \left|\frac{\mathrm{d}(Tx_1)}{\mathrm{d}t} - \frac{\mathrm{d}(Tx_2)}{\mathrm{d}t}\right|\right]^2 \mathrm{d}t \\
&\leqslant (\delta+1)^2 K^2 \int_0^\delta \mathrm{e}^{-2\lambda K t}\left[\int_0^t \mathrm{e}^{\lambda K s}\,\mathrm{e}^{-\lambda K s}\left(|x_1(s) - x_2(s)|\right.\right.
\end{aligned}$$

$$
\begin{aligned}
&+\left|\frac{\mathrm{d}x_1}{\mathrm{d}s}(s)-\frac{\mathrm{d}x_2}{\mathrm{d}s}(s)\right|\Bigg)\mathrm{d}s\Bigg]^2\,\mathrm{d}t\\
&\leqslant (\delta+1)^2K^2\int_0^\delta \mathrm{e}^{-2\lambda Kt}\left[\left(\int_0^t[\mathrm{e}^{\lambda Ks}]^2\,\mathrm{d}s\right)^{\frac{1}{2}}\right.\\
&\quad\left.\times\left(\int_0^t \mathrm{e}^{-2\lambda Ks}\left(|x_1(s)-x_2(s)|+\left|\frac{\mathrm{d}x_1}{\mathrm{d}s}(s)-\frac{\mathrm{d}x_2}{\mathrm{d}s}(s)\right|\right)^2\mathrm{d}s\right)^{\frac{1}{2}}\right]^2\mathrm{d}t\\
&=(\delta+1)^2K^2\int_0^\delta \mathrm{e}^{-2\lambda Kt}\left(\frac{\mathrm{e}^{2\lambda Kt}-1}{2\lambda K}\right)\\
&\quad\times\int_0^t \mathrm{e}^{-2\lambda Ks}\left(|x_1(s)-x_2(s)|+\left|\frac{\mathrm{d}x_1}{\mathrm{d}s}(s)-\frac{\mathrm{d}x_2}{\mathrm{d}s}(s)\right|\right)^2\mathrm{d}s\,\mathrm{d}t\\
&\leqslant\frac{(\delta+1)^2K}{2\lambda}\int_0^\delta\int_0^t \mathrm{e}^{-2\lambda Ks}\left(|x_1(s)-x_2(s)|+\left|\frac{\mathrm{d}x_1}{\mathrm{d}s}(s)-\frac{\mathrm{d}x_2}{\mathrm{d}s}(s)\right|\right)^2\mathrm{d}s\,\mathrm{d}t\\
&=\frac{(\delta+1)^2K}{2\lambda}\int_0^\delta\int_\delta^s \mathrm{e}^{-2\lambda Ks}\left(|x_1(s)-x_2(s)|+\left|\frac{\mathrm{d}x_1}{\mathrm{d}s}(s)-\frac{\mathrm{d}x_2}{\mathrm{d}s}(s)\right|\right)^2\mathrm{d}t\,\mathrm{d}s\\
&=\frac{(\delta+1)^2K}{2\lambda}\int_0^\delta(\delta-s)\,\mathrm{e}^{-2\lambda Ks}\left(|x_1(s)-x_2(s)|+\left|\frac{\mathrm{d}x_1}{\mathrm{d}s}(s)-\frac{\mathrm{d}x_2}{\mathrm{d}s}(s)\right|\right)^2\mathrm{d}s\\
&\leqslant\frac{(\delta+1)^2\delta K}{2\lambda}\|x_1-x_2\|_{2,\lambda}^2.
\end{aligned}
$$

Now pick $\lambda=(\delta+1)^2\delta K$ to conclude that

$$\|Tx_1-Tx_2\|_{2,(\delta+1)^2\delta K}\leqslant\tfrac{1}{\sqrt{2}}\|x_1-x_2\|_{2,(\delta+1)^2\delta K}.$$

□

Appendix B

One can construct a gradient descent scheme for the inverse problem in ordinary differential equations. Let us assume that the ODE is autonomous and that $f(x)$ is polynomial in x, that is

$$f(x)=\sum_{k=0}^{N}c_kx^k. \tag{33}$$

Then the fixed point $\bar{u}(t)$ of the Picard operator (7) satisfies

$$
\begin{aligned}
\bar{u}(t)&=x_0+\int_{t_0}^t f(\bar{u}(s),s)\,\mathrm{d}s\\
&=x_0+\int_{t_0}^t\sum_{k=0}^{N}c_k(\bar{u}(s))^k\,\mathrm{d}s\\
&=x_0+\sum_{k=0}^{N}c_kg_k(t),
\end{aligned}\tag{34}
$$

where $g_k(t)=\int_0^t(\bar{u}(s))^k\,\mathrm{d}s$, $k=0,\ldots,N$. Now consider $\bar{u}(t)$ to be a function of t as well as the parameters $c_1,\ldots,c_N$ and x_0. The squared $\mathcal{L}^2$ error in approximating a given target function $x(t)$ by the fixed point $\bar{u}(t)$ is $E(c_1,\ldots,c_N,x_0)=\|x-\bar{u}\|^2$, from which we calculate

$$\frac{\partial E}{\partial c_l} = -2\left\langle \frac{\partial \bar{u}}{\partial c_l}, x - \bar{u} \right\rangle, \tag{35}$$

$$\frac{\partial E}{\partial x_0} = -2\left\langle \frac{\partial \bar{u}}{\partial x_0}, x - \bar{u} \right\rangle. \tag{36}$$

Hence we find the partial derivative of $\bar{u}(t)$ with respect to each parameter. We have

$$\begin{aligned}\frac{\partial \bar{u}}{\partial c_l} &= g_l + \sum_{k=0}^{N} c_k \frac{\partial g_k}{\partial c_l} \\ &= g_l + \sum_{k=1}^{N} k c_k \int_0^t (\bar{u}(s))^{k-1} \frac{\partial \bar{u}}{\partial c_l}(s)\,\mathrm{d}s,\end{aligned}$$

which upon differentiation with respect to t gives

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \bar{u}}{\partial c_l}\right) &= \frac{\mathrm{d}g_l}{\mathrm{d}t} + \sum_{k=1}^{N} k c_k\,(\bar{u}(t))^{k-1}\,\frac{\partial \bar{u}}{\partial c_l}(t), \\ &= (\bar{u}(t))^k + \sum_{k=1}^{N} k c_k\,(\bar{u}(t))^{k-1}\,\frac{\partial \bar{u}}{\partial c_l}(t).\end{aligned} \tag{37}$$

Equation (37) is a first-order linear DE with solution

$$\frac{\partial \bar{u}}{\partial c_l} = \int_0^t (\bar{u}(s))^l \exp\left(\sum_{k=1}^{N} k c_k (g_{k-1}(t) - g_{k-1}(s))\,\mathrm{d}s\right). \tag{38}$$

Now differentiate (34) with respect to x_0 to get

$$\begin{aligned}\frac{\partial \bar{u}}{\partial x_0} &= 1 + \sum_{k=0}^{N} c_k \frac{\partial g_k}{\partial x_0} \\ &= 1 + \sum_{k=1}^{N} k c_k \int_0^t (\bar{u}(s))^{k-1} \frac{\partial \bar{u}}{\partial x_0}(s)\,\mathrm{d}s,\end{aligned}$$

which upon differentiation with respect to t gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \bar{u}}{\partial x_0}\right) = \sum_{k=1}^{N} k c_k (\bar{u}(t))^{k-1} \frac{\partial \bar{u}}{\partial x_0}(t). \tag{39}$$

The solution to (39) is

$$\frac{\partial \bar{u}}{\partial x_0} = \exp\left(\sum_{k=1}^{N} k c_k g_{k-1}(t)\right). \tag{40}$$

A gradient descent scheme now follows by putting together (35), (36), (38) and (40). The computational difficulty is that the construction of the gradient vector requires knowledge of the solution at the current parameter values. This solution is likely only known numerically, which means that the various integrals involving it will have to be calculated numerically, potentially breeding non-negligible numerical error.

References

[1] Barnsley M, Ervin V, Hardin D and Lancaster J 1985 Solution of an inverse problem for fractals and other sets *Proc. Natl Acad. Sci. USA* **83** 1975–7

[2] Bellman R 1966 Inverse problems in ecology *J. Theor. Biol.* **11** 164–7

[3] Bielecki A 1956 Une remarque sur la méthode de Banach–Caccioppoli–Tikhonov dans la théorie des équations differentielles ordinaires *Bull. Acad. Pol. Sci. Cl. III* **4** 261–4

[4] Centore P and Vrscay E R 1994 Continuity of attractors and invariant measures for iterated function systems *Can. Math. Bull.* **37** 315–29

[5] Coddington E A and Levinson N 1955 *Theory of Ordinary Differential Equations* (New York: McGraw-Hill)

[6] Crutchfield J and McNamara B 1987 Equations of motion from a data series *Complex Syst.* **1** 417–52

[7] Dudbridge F and Fisher Y 1997 Attractor optimization in fractal image coding *Proc. 3rd Fractals in Engineering Conf. (Arcachon, France, May 1997)*

[8] Fisher Y 1995 *Fractal Image Compression, Theory and Application* (New York: Springer)

[9] Fisher Y (ed) 1998 *Fractal Image Coding and Analysis, Proc. NATO ASI* (New York: Springer)

[10] Forte B and Vrscay E R 1998 Inverse problem methods for generalized fractal transforms *Fractal Image Encoding and Analysis (NATO ASI Series F* vol 159*)* ed Y Fisher (Heidelberg: Springer)

[11] Groetsch C W 1993 *Inverse Problems in the Mathematical Sciences* (Braunschweig: Vieweg)

[12] Himmelblau D M, Jones C R and Bischoff K B 1967 Determination of rate constants for complex kinetics models *Ind. Eng. Chem. Fundam.* **6** 539–43

[13] Kunze H and Vrscay E 1999 Solving inverse problems for ODEs using the Picard contraction mapping *Inverse Probl.* **15** 745–70

[14] Lanczos C 1956 *Applied Analysis* (Englewood Cliffs, NJ: Prentice-Hall) pp 272–80

[15] Lu N 1997 *Fractal Imaging* (New York: Academic)

[16] Marsili-Libelli S 1992 Parameter estimation of ecological models *Ecol. Model.* **62** 233–58

[17] Milstein J 1981 The inverse problem: estimation of kinetic parameters *Modeling of Chemical Reaction Systems* ed K H Ebert, P Deuflhard and W Jäger (Berlin: Springer)

[18] Ruhl M and Hartenstein H 1997 Optimal fractal coding is NP-hard *Proc. IEEE Data Compression Conf. (Snowbird, Utah, 1997)* ed J Storer and M Cohn

[19] Tanner R 1972 Estimating kinetic rate constants using orthogonal polynomials and Picard's iteration method *Ind. Eng. Chem. Fundam.* **11** 1–8

[20] Vrscay E R and Saupe D 1999 Can one break the 'collage barrier' in fractal image coding *Fractals: Theory and Applications in Engineering* ed M Dekking, J Levy-Vehel, E Lutton and C Tricot (Berlin: Springer) pp 307–23