Applied Mathematics Letters 25 (2012) 2306–2311

Contents lists available at SciVerse ScienceDirect

Applied Mathematics Letters

journal homepage: www.elsevier.com/locate/aml

Solving inverse problems for DEs using the Collage Theorem and entropy maximization

Herb Kunze [a], Davide La Torre [b,*], Edward R. Vrscay [c]

[a] University of Guelph, Canada
[b] University of Milan, Milan, Italy
[c] University of Waterloo, Ontario, Canada

ARTICLE INFO

Article history:
Received 24 January 2012
Received in revised form 2 June 2012
Accepted 17 June 2012

Keywords:
Inverse problems for DEs
Collage Theorem
Entropy

ABSTRACT

In this work, we focus on the inverse problem associated with a DE: Given a target function x, find a DE such that its solution $\bar{x}$ is sufficiently close to x in the sup norm distance. We extend the previous method for solving inverse problems for DEs using the Collage Theorem along a new direction. We search for a set of coefficients that not only minimizes the collage error but also maximizes the entropy. This approach is motivated by some promising results for IFS and probabilities. In our new formulation, the minimization of the collage error can be understood as a multi-criteria problem: two different and conflicting criteria are considered, i.e., collage error and the entropy. In order to deal with this kind of scenario we propose to *scalarize* the model, which reduces the multi-criteria program to a single-criterion program by combining all objective functions with different trade-off weights. Numerical examples confirm the sub-optimality of the Collage Theorem and we show that, by adding the entropy term, we obtain a better approximation of the solution.

1. Inverse problems for fixed point equations

Many inverse problems may be viewed in terms of the approximation of a target element x in a complete metric space (X, d) by the fixed point $\bar{x}$ of a contraction mapping $T : X \to X$. In practical applications, from a family of contraction mappings $T_\lambda,\ \lambda \in \Lambda \subset \mathbb{R}^n$, one wishes to find the parameter $\bar{\lambda}$ for which the approximation error $d(x, \bar{x}_\lambda)$ is as small as possible. Thanks to a simple consequence of Banach's fixed point theorem known as the "Collage Theorem", most practical methods of solving the inverse problem for fixed point equations seek to find an operator T for which the *collage distance* $d(x, Tx)$ is as small as possible.

Theorem 1 (*"Collage Theorem" [1]*). *Let (X, d) be a complete metric space and $T : X \to X$ a contraction mapping with contraction factor $c \in [0, 1)$. Then for any $x \in X$,*

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

where $\bar{x}$ is the fixed point of T.

One now seeks a contraction mapping T that minimizes the so-called *collage error* $d(x, Tx)$—in other words, a mapping that sends the target x as close as possible to itself. This is the essence of the method of *collage coding* which has been the basis of most, if not all, fractal image coding and compression methods.

* Corresponding author.
E-mail address: davide.latorre@unimi.it (D. La Torre).

0893-9659/$ – see front matter
doi:10.1016/j.aml.2012.06.021

2. Inverse problems for DEs investigated using the Collage Theorem

In [2–5], the authors showed how collage coding could be used to solve inverse problems for systems of differential equations (deterministic or random) having the form

$$\begin{cases} \dot{x}(t) = f(t, x), \\ x(0) = x_0, \end{cases} \tag{2}$$

by reducing the problem to the corresponding Picard integral operator associated with it,

$$(Tx)(t) = x_0 + \int_0^t f(s, x(s))\, ds. \tag{3}$$

Let us recall the basic results in the case where f belongs to L^2. Let us consider the complete metric space $C([-\delta, \delta])$ endowed with the usual d_∞ metric and assume that $f(t, x)$ is Lipschitz in the variable x, that is there exists a $K \geq 0$ such that $|f(s, x_1) - f(s, x_2)| \leq K|x_1 - x_2|$, for all $x_1, x_2 \in \mathbb{R}$. For simplicity we suppose that $x \in \mathbb{R}$ but the same consideration can be developed for the case of several variables. Under these hypotheses, T is Lipschitz on the space $C([-\delta, \delta] \times [-M, M])$ for some δ and $M > 0$.

Theorem 2 ([2]). *The function T satisfies*

$$\|Tx - Ty\|_2 \leq c\|x - y\|_2 \tag{4}$$

for all $x, y \in C([-\delta, \delta] \times [-M, M])$ where $c = \delta K$.

Now let $\delta' > 0$ be such that $\delta' K < 1$. In order to solve the inverse problem for (3) we take the L^2 expansion of the function f. Let $\{\phi_i\}$ be a basis of functions in $L^2([-\delta', \delta'] \times [-M, M])$ and consider

$$f_\lambda(s, x) = \sum_{i=1}^{+\infty} \lambda_i \phi_i(s, x). \tag{5}$$

Each sequence of coefficients $\lambda = \{\lambda_i\}_{i=1}^{+\infty}$ then defines a Picard operator T_λ. Suppose further that each function $\phi_i(s, x)$ is Lipschitz in x with constants K_i.

Theorem 3 ([2]). *Let $K, \lambda \in \ell^2(\mathbb{R})$. Then*

$$|f_\lambda(s, x_1) - f_\lambda(s, x_2)| \leq \|K\|_2 \|\lambda\|_2 |x_1 - x_2| \tag{6}$$

for all $s \in [-\delta', \delta']$ and $x_1, x_2 \in [-M, M]$ where $\|K\|_2 = \left(\sum_{i=1}^{+\infty} K_i^2\right)^{\frac{1}{2}}$ and $\|\lambda\|_2 = \left(\sum_{i=1}^{+\infty} \lambda_i^2\right)^{\frac{1}{2}}$.

Given a target solution x, we now seek to minimize the collage distance $\|x - T_\lambda x\|_2$. The square of the collage distance becomes

$$\begin{aligned} \Delta^2(\lambda) &= \|x - T_\lambda x\|_2^2 \\ &= \int_{-\delta}^{\delta} \left| x(t) - x_0 - \int_0^t \sum_{i=1}^{+\infty} \lambda_i \phi_i(s, x(s)) ds \right|^2 dt \end{aligned} \tag{7}$$

and the inverse problem can be formulated as

$$\min_{\lambda \in \Lambda} \Delta(\lambda), \tag{8}$$

where $\Lambda = \{\lambda \in \ell^2(\mathbb{R}) : \|\lambda\|_2 \|K\|_2 < 1\}$. To solve this problem numerically, let us consider the first n terms of the L^2 basis; in this case the previous problem can be reduced to

$$\min_{\lambda \in \tilde{\Lambda}} \tilde{\Delta}^2(\lambda) = \int_{-\delta}^{\delta} \left| x(t) - x_0 - \int_0^t \sum_{i=1}^{n} \lambda_i \phi_i(s, x(s)) ds \right|^2 dt, \tag{9}$$

where $\tilde{\Lambda} = \{\lambda \in \mathbb{R}^n : \|\lambda\|_2 \|K\|_2 < 1\}$. This is a classical quadratic optimization problem which can be solved by means of classical numerical methods. Let $\tilde{\Delta}^n_{\min}$ be the minimum value of $\tilde{\Delta}$ over $\tilde{\Lambda}$. This is a non-increasing sequence of numbers (depending on n) and as shown in [6] it is possible to show that $\liminf_{n \to +\infty} \tilde{\Delta}^n_{\min} = 0$. This states that the distance between the target element and the unknown solution of the differential equation can be made arbitrary small.

For further discussion of the collage method and much related analysis, the reader is referred to [7].

3. Collage error and entropy maximization

In this section, we introduce a second criterion for selecting the unknown parameters when solving the inverse problem for DEs and this criterion incorporates a maximum entropy philosophy [8,9]. It is well known that in Bayesian probability the principle of maximum entropy is an axiom, stating that *the probability distribution which best represents the current state of knowledge is the one with largest information theoretical entropy*. This approach has already been successfully applied to inverse problems for iterated function systems with probabilities [10]. The idea behind our approach is to improve the approximation of the target by using a scalarization procedure to combine the criteria of maximum entropy and collage distance minimization. The usual Shannon entropy term takes the form $p_i \ln(p_i)$, where $0 < p_i < 1$; we adapt each unknown parameter λ_i to this framework by constructing an entropy term of the form $\left|\frac{\lambda_i}{M}\right| \ln \left|\frac{\lambda_i}{M}\right|$, where the value of $M > 0$ is chosen such that $\left|\frac{\lambda_i}{M}\right| < 1$.

Thus, we consider a multi-criteria optimization problem which involves the following two objective functions:

- $f_1(\lambda) = \tilde{\Delta}^2(\lambda) = \int_{-\delta}^{\delta} \left| x(t) - x_0 - \int_0^t \sum_{i=1}^n \lambda_i \phi_i(s, x(s)) ds \right|^2 dt$
- $f_2(\lambda) = \sum_{i=1}^n \left|\frac{\lambda_i}{M}\right| \ln \left|\frac{\lambda_i}{M}\right|$

where M is such that $\tilde{\Lambda} = \{\lambda \in \mathbb{R}^n : \|\lambda\|_2 \|K\|_2 \le C < 1\} \subset B_M(0)$ with $B_M(0)$ being the ball centered at the origin with radius M and C a fixed positive number strictly less than 1. The first criterion has been well explained in the previous paragraph and expresses the collage distance between the target and the Picard operator applied to the target in terms of the expansion. The second one is the negative of the classical definition of Shannon entropy although, of course, other alternatives for the entropy functional could be considered. It seems quite natural to combine these two criteria in a unique framework by scalarizing the multi-criteria problem through a trade-off weight $\eta \in [0, 1]$. This leads to the following model:

$$\min f_1(\lambda) + \eta f_2(\lambda) \tag{10}$$

subject to $\lambda \in \tilde{\Lambda}$. The following Proposition 1 illustrates which kinds of relationships there are between optimal solutions to Eq. (10) and optimal solutions to the collage minimization problem

$$\min f_1(\lambda) \tag{11}$$

which corresponds to $\eta = 0$. Since $\tilde{\Lambda}$ is a compact set and f_1 is a continuous function, then both Eqs. (11) and (10) admit a global solution (the negative entropy f_2 is a continuous strict convex function). Let λ_0 be an optimal solution to Eq. (11) and λ_η be an optimal solution to Eq. (10) for any fixed $\eta \in [0, 1]$.

Proposition 1. *The negative entropy obtains its maximum at* λ_0.

Proof. Since the feasible set is invariant, it is easy to prove that

$$f_1(\lambda_0) \le f_1(\lambda_\eta) \tag{12}$$

and

$$f_1(\lambda_\eta) + \eta f_2(\lambda_\eta) \le f_1(\lambda_0) + \eta f_2(\lambda_0) \le f_1(\lambda_\eta) + \eta f_2(\lambda_0). \tag{13}$$

The first and the last term of this chain of inequalities imply that

$$f_2(\lambda_\eta) \le f_2(\lambda_0). \quad \square \tag{14}$$

Recalling that f_2 is the negative entropy, according to the maximum entropy principle, it looks reasonable to search for solutions to Eq. (10) which provide better approximations than the classical collage minimization approach. In the Appendix we provide a sufficient condition which analytically demonstrates that the approximation we obtain by adding the entropy is better than the one we have via the classical collage approach. In the following section we list some numerical simulations which show how the method works.

3.1. Numerical simulations

We present two examples which illustrate that allowing the entropy criterion to play a small role in the scalarized problem (10) improves the error in the resulting approximation.

Example 1. In this numerical example we use the simple linear initial value problem

$$\dot{x}(t) = 2 + 3x, \qquad x(0) = 0.5,$$

with solution

$$x(t) = -\frac{2}{3} + \frac{7}{6} e^{3t}.$$

Table 1
Results for Example 1 for different choices of η.

η	$\Delta(\lambda)$	$f_2(\lambda)$	ER
0.000	0.0006982	−0.5391862	0.0172267
0.010	0.0006997	−0.5394967	0.0167919
0.013	0.0007008	−0.5395901	0.0167542
0.014	0.0007012	−0.5396213	0.0167519

Table 2
Results for Example 2 for different choices of η.

η	$\Delta(\lambda)$	$f_2(\lambda)$	ER
0.0000	0.0008874	−0.3320031	0.0175538
0.0010	0.0008875	−0.3320320	0.0175423
0.0015	0.0008875	−0.3320464	0.0175458
0.0019	0.0008875	−0.3320580	0.0175532

We sample the solution at 21 uniformly spaced times in [0, 1], add normally distributed noise, and fit an eighth-degree polynomial to the 21 data points, calling the resulting polynomial $x_{\text{target}}(t)$. The inverse problem is: Find an initial value problem $\dot{x}(t) = g^{\lambda}(x) = \lambda^0 + \lambda^1 x, x(0) = x^0$, that admits $x_{\text{target}}(t)$ as an approximate solution. The parameters of the problem are $\lambda = (x^0, \lambda^0, \lambda^1)$. The collage distance criterion is

$$f_1(\lambda) = \tilde{\Delta}^2(\lambda) = \int_0^1 \left| x_{\text{target}}(t) - x^0 - \int_0^t \left(\lambda^0 + \lambda^1 x_{\text{target}}(s)\right) ds \right|^2 dt.$$

When we set $M = 25$, the entropy criterion is

$$f_2(\lambda) = \left|\frac{x^0}{25}\right| \ln\left|\frac{x^0}{25}\right| + \left|\frac{\lambda^0}{25}\right| \ln\left|\frac{\lambda^0}{25}\right| + \left|\frac{\lambda^1}{25}\right| \ln\left|\frac{\lambda^1}{25}\right|.$$

For chosen η, we seek to solve

$$\min_{x^0,\lambda^0,\lambda^1} f_1(\lambda) + \eta f_2(\lambda).$$

We solve this nonlinear optimization problem by using the nonlinear program solver in Maple, `NLPSolve`. Finally, we use the values obtained for x^0, λ^0, and λ^1, solving the initial value problem $\dot{y}(t) = g^{\lambda}(x)$, $y(0) = x^0$, and measuring the L^2 error

$$ER = d_2(x, y) = \left(\int_0^1 (x(t) - y(t))^2\, dt\right)^{\frac{1}{2}}.$$

The results are presented in Table 1. The first row of the table presents the case of minimizing the squared collage distance only. The subsequent rows of the table give results for other choices of η, allowing the entropy criterion to play a role. We see that the collage distance is higher in these rows, but that the true error in the L^2 approximation of $x(t)$ by $y(t)$ is smaller.

Example 2. We repeat the procedure of Example 1, this time with the quadratic initial value problem

$$\dot{x}(t) = x(1 - x), \qquad x(0) = -0.5,$$

with solution

$$x(t) = \frac{1}{1 - e^{-3t}}.$$

We construct $x_{\text{target}}(t)$ in the same manner and consider the same inverse problem, this time with $g^{\lambda}(x) = \lambda^1 x + \lambda^2 x^2$. The parameters of the problem are $\lambda = (x^0, \lambda^1, \lambda^2)$. We construct the collage distance and the entropy, again using $M = 25$ in the latter, arriving at the problem

$$\min_{x^0,\lambda^1,\lambda^2} f_1(\lambda) + \eta f_2(\lambda)$$

for chosen η. Using Maple's `NLPSolve`, we obtain parameter values, for which we solve the initial value problem $\dot{y}(t) = g^{\lambda}(x)$, $y(0) = x^0$, so that we can measure the L^2 error ER as in Example 1.

The results are presented in Table 2. The first row of the table presents the case of minimizing the squared collage distance only, and the subsequent rows of the table give the entropy criterion some weight. Once again, we see that the collage distance is higher in these latter rows, but that the true error in the L^2 approximation of $x(t)$ by $y(t)$ is smaller.

4. Conclusions

The calculations performed show that a kind of maximum entropy principle holds for inverse problems for DEs. We have observed that the addition of an entropy term into the optimization problem yields better approximations to the true solution without any reduction of the collage error.

Appendix

Here we again consider the classical collage minimization problem,

$$\min f_1(\lambda), \tag{15}$$

and the collage minimization problem with entropy,

$$\min f_1(\lambda) + \eta f_2(\lambda), \tag{16}$$

both subject to $\lambda \in \tilde{\Lambda} = \{\lambda \in \mathbb{R}^n : \|\lambda\|_2 \|K\|_2 \leq C < 1\}$ and $\eta \in [0, 1]$. Let λ_0 and λ_η be two optimal solutions to problems (15) and (16), respectively.

Now consider the following function $\phi(\lambda)$ which measures an appropriately weighted distance between the squared collage error $f_1(\lambda)$ and the squared approximation error $\|u - u_\lambda\|^2$:

$$\phi(\lambda) := f_1(\lambda) - (1 - C)^2 \|u - u_\lambda\|^2. \tag{17}$$

Note that the Collage Theorem (Section 1) implies that $\phi(\lambda) \geq 0$ for any $\lambda \in \Lambda$.

We now show that if $\phi(\lambda)$ satisfies two appropriate conditions, then the approximation error associated with Problem (16) is *lower* than its counterpart for Problem (15), i.e.,

$$\|u - u_{\lambda_\eta}\| \leq \|u - u_{\lambda_0}\|. \tag{18}$$

The conditions to be satisfied by $\phi(\lambda)$ in this discussion are as follows:

1. $\phi(\lambda_0) = 0$.
2. There exists an $\eta > 0$ such that $\phi(\lambda_\eta) \geq \eta[f_2(\lambda_0) - f_2(\lambda_\eta)] \geq 0$.

A couple of remarks regarding these conditions are in order.

1. Condition 1 implies that the minimum possible value for the collage error Δ for Problem (15) has been achieved, according to the Collage Theorem.
2. Perhaps a better understanding of Condition 2 is achieved if it is rewritten as follows, using the fact that $\phi(\lambda_0) = 0$:

$$\phi(\lambda_\eta) - \phi(\lambda_0) \leq (-\eta)\left[f_2(\lambda_\eta) - f_2(\lambda_0)\right] \geq 0. \tag{19}$$

In other words, the net change in $\phi(\lambda)$ is related to the net change in $f_2(\lambda)$ in a rather special way, namely, a negative proportionality. (From Eq. (14) in the main text, the quantity in the square brackets is negative.) Although this condition is difficult to check because it requires the knowledge of λ_η – which is unknown – it is reasonable to think that this condition may satisfied for small values of η. (Trivially, $\eta[f_2(\lambda_0) - f_2(\lambda_\eta)] \to 0$ as $\eta \to 0$.)

We now proceed with the derivation. The second condition on $\phi(\lambda)$ implies that

$$f_1(\lambda_\eta) - (1 - C)^2 \|u - u_{\lambda_\eta}\|^2 \geq \eta[f_2(\lambda_0) - f_2(\lambda_\eta)], \tag{20}$$

which may be rearranged to yield

$$\eta f_2(\lambda_\eta) + f_1(\lambda_\eta) - (1 - C)^2 \|u - u_{\lambda_\eta}\|^2 \geq \eta f_2(\lambda_0). \tag{21}$$

Since λ_η is a global minimum of (16), it follows that

$$\eta f_2(\lambda_0) + f_1(\lambda_0) - (1 - C)^2 \|u - u_{\lambda_\eta}\|^2 \geq \eta f_2(\lambda_0). \tag{22}$$

This result implies that

$$f_1(\lambda_0) = (1 - C)^2 \|u - u_{\lambda_0}\|^2 \geq (1 - C)^2 \|u - u_{\lambda_\eta}\|^2, \tag{23}$$

which yields the desired result in Eq. (18).

In summary, we have shown conditions on the "distance function" $\phi(\lambda)$ that imply that the approximation u_{λ_η} obtained by a minimization model with entropy, i.e., Eq. (16) is better than the approximation u_{λ_0} obtained from the classical minimization of the collage distance.

References

[1] M.F. Barnsley, Fractals Everywhere, Academic Press, New York, 1989.
[2] H. Kunze, E.R. Vrscay, Solving inverse problems for ordinary differential equations using the Picard contraction mapping, Inverse Problems 15 (1999) 745–770.
[3] H. Kunze, J. Hicken, E.R. Vrscay, Inverse problems for ODEs using contraction maps: suboptimality of the collage method, Inverse Problems 20 (2004) 977–991.
[4] H. Kunze, D. La Torre, E.R. Vrscay, Inverse problems for random differential equations using the collage method for random contraction mappings, Journal of Computational and Applied Mathematics 223 (2) (2009) 853–861.
[5] H. Kunze, D. La Torre, E.R. Vrscay, Random fixed point equations and inverse problems using the collage method for contraction mappings, Journal of Mathematical Analysis and Applications 334 (2) (2007) 1116–1129.
[6] B. Forte, E.R. Vrscay, Inverse problem methods for generalized fractal transforms, in: Y. Fisher (Ed.), Fractal Image Encoding and Analysis, in: NATO ASI Series F, vol. 159, Springer-Verlag, New York, 1998.
[7] H. Kunze, D. La Torre, F. Mendivil, E.R. Vrscay, Fractal-Based Methods in Analysis, Springer, 2012.
[8] S.F. Gull, J. Skilling, Maximum entropy method in image processing, Proceedings of IEEE, Part F 131 (6) (1984) 646–659.
[9] E.T. Jaynes, in: G.L. Bretthorst (Ed.), Probability Theory: The Logic of Science, Cambridge University Press, Cambridge, 2003.
[10] D. La Torre, E.R. Vrscay, Fractal-based measure approximation with entropy maximization and sparsity constraints, AIP Conference Proceedings 1443 (2012) 63–71.