COLLAGE-BASED INVERSE PROBLEMS FOR IFSM WITH ENTROPY MAXIMIZATION AND SPARSITY CONSTRAINTS

HERB KUNZE✉,1, DAVIDE LA TORRE² AND EDWARD VRSCAY³

¹Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada; ²Department of Applied Mathematics and Sciences, Khalifa University, Abu Dhabi, UAE and Department of Economics, Management and Quantitative Methods, University of Milan, Italy; ³Department of Applied Mathematics, University of Waterloo, Ontario, Canada
e-mail: hkunze@uoguelph.ca, davide.latorre@unimi.it, ervrscay@uwaterloo.ca
(Received July 8, 2013; revised November 1, 2013; accepted November 1, 2013)

ABSTRACT

We consider the inverse problem associated with IFSM: Given a target function f, find an IFSM, such that its invariant fixed point $\bar{f}$ is sufficiently close to f in the L^p distance. In this paper, we extend the collage-based method developed by Forte and Vrscay (1995) along two different directions. We first search for a set of mappings that not only minimizes the collage error but also maximizes the entropy of the dynamical system. We then include an extra term in the minimization process which takes into account the sparsity of the set of mappings. In this new formulation, the minimization of collage error is treated as multi-criteria problem: we consider three different and conflicting criteria *i.e.*, collage error, entropy and sparsity. To solve this multi-criteria program we proceed by scalarization and we reduce the model to a single-criterion program by combining all objective functions with different trade-off weights. The results of some numerical computations are presented. Numerical studies indicate that a maximum entropy principle exists for this approximation problem, *i.e.*, that the suboptimal solutions produced by collage coding can be improved at least slightly by adding a maximum entropy criterion.

Keywords: collage theorem, entropy, fractal transforms, iterated function systems with mappings, sparsity.

INTRODUCTION

In fractal image coding based on Generalized Fractal Transforms (GFT), one seeks to approximate a target image or signal v by the fixed point $\bar{u}$ of a contractive fractal transform operator T. The usual formulation involves a fixed set of geometric contraction maps along with a corresponding set of greyscale maps. The inverse problem, which involves the determination of the best greyscale map parameters for a given target image, is based on the so-called "Collage Theorem", a simple consequence of Banach's fixed point theorem. Another consequence of Banach's fixed point result is that the approximation of the target image or signal can be generated by iteration of the fractal transform (see Hutchinson, 1981; Barnsley et al., 1985; Barnsley and Demko, 1985; Barnsley, 1989; Barnsley and Hurd, 1993; Forte and Vrscay, 1995; 1999; Iacus and La Torre, 2005a; 2005b; Kunze et al., 2008; 2009; 2012a; 2012b; La Torre et al., 2009; La Torre and Vrscay, 2009; 2011). In Forte and Vrscay (1995), the authors showed that one can find an iterated function system with greyscale maps (IFSM) to approximate any target signal or image with arbitrary precision, and they provided a suboptimal but systematic "collage-based" approach for doing so.

In this paper we extend the approach developed

in Forte and Vrscay (1995) along two different – in fact, competing – directions, namely, entropy and sparsity. First, we search for a set of mappings and greyscale map parameters that not only minimizes the so-called *collage error* but also maximizes the entropy of the parameter set. A motivation for this procedure is as follows. As stated earlier, given a target image v, one would ideally like to find a contractive fractal transform T with fixed point $\bar{u}$ that is as close as possible to v, *i.e.*, which minimizes the approximation error given by the distance $E(u) = d(v,u)$. This problem, however, is enormously difficult and impractical. The Collage Theorem provides a significant simplification in that one searches for a fractal transform T – as defined by its fractal parameters – which minimizes the "collage distance" $d(v,Tv)$. Such a procedure is often easy to formulate and solve algorithmically. As may be expected, however, the fractal transform T_c yielded by this collage-based method is suboptimal, *i.e.*, its fixed point $\bar{u}_c$ does not minimize the true approximation error $E(u)$. As such, we may consider the parameters defining the suboptimal fractal transform T_c as representing "incomplete" or "partial information." We now apply the maximum entropy principle of Jaynes (1957), *i.e.*, "making inferences on the basis of partial information we must use that probability distribution

which has maximum entropy subject to whatever is known." Here, a suitable probability distribution over the relevant fractal parameters is defined and employed.

Second, we examine the effects of maximizing the *sparsity* of the set of greyscale parameters, *i.e.*, forcing as many of these parameters as possible to be zero. The motivation of this approach is as follows. The parameters defining the suboptimal transform T_c are once again viewed as representing an approximation to the true solution. We now borrow from sparsity studies in signal and image processing, where it is often found that a signal/image is well approximated by a vector of coefficients with a small number of nonzeros (see Elad, 2010) – in other words, a vector with *high sparsity*.

In the new formulation presented in this paper, the minimization of collage error is studied as a multi-criteria problem. Three different and conflicting criteria are considered, namely collage error, entropy and sparsity. In order to reduce the complexity of this model, we employ a scalarization technique which allows the multi-criteria program to be reduced to a single-criterion program by combining all objective functions with different trade-off weights. This new approach is illustrated through some numerical examples which indicate that a maximum entropy principle does exist for this approximation problem, *i.e.*, that the suboptimal solutions produced by collage coding can be improved at least slightly by adding a maximum entropy criterion. Similar results were obtained in the context of measure approximation using Iterated Function System with Probabilities (IFSP) by La Torre and Vrscay (2012) and in the analysis of inverse problems for differential equations by collage methods in Kunze et al. (2012).

ITERATED FUNCTION SYSTEMS ON FUNCTIONS: SOME BASICS

The action of a GFT $T : X \to X$ on an element u of the complete metric space (X, d) can be summarized in the following steps. It produces a set of N spatially-contracted copies of u and then it modifies the values of these copies by means of a suitable range-mapping. Finally, it recombines them using an appropriate operator in order to get the element $v \in X$, $v = Tu$. In all these cases, under appropriate conditions, the fractal transform T is a contraction and thus Banach's fixed point theorem guarantees the existence of a unique fixed point $\bar{u} = T\bar{u}$. The inverse problem is a key factor for applications: given $T : X \to X$ a point-to-point contraction mapping and a "target" element

$v \in X$, we look for a contraction mapping T with fixed point $\bar{u}$ such that $d(v, \bar{u})$ is as small as possible. In practical applications, however, it is difficult to construct solutions to this problem and one relies on the following simple consequence of Banach's fixed point theorem, known in the fractal coding literature as the *collage theorem*, which states that

$$d(v, \bar{u}) \leq \frac{1}{1-c} d(v, Tv) \tag{1}$$

(c is the contractivity factor of T). Instead of trying to minimize the error $d(v, \bar{u})$, one looks for a contraction mapping T that minimizes the *collage error $d(v, Tv)$*. In this section we focus on the method of iterated function systems with greyscale maps (IFSM), as formulated by Forte and Vrscay (1995), a GFT which can be used to approximate a given element u of $L^2([0, 1])$.

We consider the case in which $u : [0, 1] \to [0, 1]$ and the space

$$X = \left\{ u : [0, 1] \to [0, 1], u \in L^2[0, 1] \right\} . \tag{2}$$

The ingredients of an *N*-map IFSM on X are

1. a set of N contractive mappings $w = \{w_1, w_2, \ldots, w_N\}$, $w_i(x) : [0, 1] \to [0, 1]$, most often affine in form:

$$w_i(x) = s_i x + a_i , \quad 0 \leq s_i < 1, \quad i = 1, 2, \ldots, N; \tag{3}$$

2. a set of associated functions—the greyscale maps—$\phi = \{\phi_1, \phi_2, \ldots, \phi_N\}$, $\phi_i : \mathbb{R} \to \mathbb{R}$. Affine maps are usually employed:

$$\phi_i(t) = \alpha_i t + \beta_i , \tag{4}$$

with the conditions

$$\alpha_i, \beta_i \in [0, 1] \tag{5}$$

and

$$0 \leq \sum_{i=1}^{N} \alpha_i + \beta_i < 1 . \tag{6}$$

Associated with the *N*-map IFSM (w, ϕ) is the *fractal transform* operator T, the action of which on a function $u \in X$ is given by

$$(Tu)(x) = \sum_{i=1}^{N} {}' \phi_i(u(w_i^{-1}(x))) , \tag{7}$$

where the prime means that the sum operates on all those terms for which w_i^{-1} is defined.

Theorem 1. *(Forte and Vrscay, 1995) $T : X \to X$ and for any $u, v \in X$ we have*

$$d_2(Tu, Tv) \le C d_2(u, v) , \qquad (8)$$

where

$$C = \sum_{i=1}^{N} s_i^{\frac{1}{2}} \alpha_i . \qquad (9)$$

When $C < 1$, then T is contractive on X, implying the existence of a unique fixed point $\bar{u} \in X$ such that $\bar{u} = T\bar{u}$.

The inverse problem associated with IFSM can, in principle, be solved to arbitrary accuracy, using a procedure defined in Forte and Vrscay (1995). The squared collage distance function associated with an N-map IFSM may be written as a quadratic form,

$$\Delta^2 = z^T A z + b^T z + c , \qquad (10)$$

where $z = (\alpha_1, \ldots \alpha_N, \beta_1, \ldots, \beta_N)$. The maps w_k are chosen from an infinite set W of fixed affine contraction maps on $[0, 1]$ which satisfy the following properties.

Definition 2.1. *We say that W generates an m-dense and nonoverlapping family A of subsets of I if for every $\varepsilon > 0$ and every $B \subset I$ there exists a finite set of integers i_k, $i_k \ge 1$, $1 \le k \le N$, such that*

- $A = \cup_{k=1}^{N} w_{i_k}(I) \subset B,$
- $m(B \backslash A) < \varepsilon$, and
- $m(w_{i_k}(I) \cap w_{i_l}(I)) = 0$ if $k \ne l,$

where m denotes Lebesgue measure.

Let
$$W^N = \{w_1, \ldots w_N\} \qquad (11)$$

be the N truncations of w. Let $\Phi^N = \{\phi_1, \ldots, \phi_N\}$ be the N-vector of affine grey level maps. Let Ω be a compact subset of set $\mathbb{R}^{2N}$ which describes the set of all possible constraints and let z_N be the solution of the previous quadratic optimization problem over Ω. Let $\Delta^2_{N,min} = \Delta^2_N(z_N)$. In Forte and Vrscay (1995), the following result was proved.

Theorem 2.

$$\Delta^2_{N,min} \to 0 \ as \ N \to \infty .$$

Using the Collage Theorem, the inverse problem may be solved to arbitrary accuracy. A practical choice for the contraction maps w on $X = [0, 1]$ is

$$w_{ij}(x) = 2^{-i}(x + j - 1), \ i = 1, 2, \ldots, \ j = 1, 2, \ldots, 2^i.$$

COLLAGE ERROR MINIMIZATION, ENTROPY AND SPARSITY MAXIMIZATION

Solutions to the optimization problem,

$$\min_{z \in \Omega} \Delta_2(z) \qquad (12)$$

using a quadratic programming algorithm were presented in Forte and Vrscay (1995). Here we consider an extension of the above optimization problem which includes two additional objective functions, namely, maximum entropy and sparsity. The reasons for adding these two criteria were discussed in the Introduction but we recall them briefly here.

On the one hand, it is desired to improve the results obtained from the collage-based method, acknowledged as being suboptimal, thereby obtaining a better approximation of the target image. In fact, from Eq. (1), we see that the collage error $d(v, Tv)$ provides an upper bound to the error, $d(v, \bar{u})$, in approximating the target v with the fixed point $\bar{u}$ of T. We consider the maximum entropy principle as a way of improving the collage-based approximation.

On the other hand, collage error minimization often provides a solution with several nonzero coefficients, many of them being very close to zero. Once again borrowing the idea of sparsity from signal/image processing, it seems to be quite natural to ask – especially in the context of data compression – if there exists an alternative solution which possesses a lesser number of nonzero coefficients – hence greater sparsity – but which still produces a small approximation error. This is the role played by the function F_2 below which counts the number of nonzero coefficients z_i – essentially the l_0 norm of the vector z – which is a measure of the *lack of sparsity* of z. Bearing these motivations in mind, the collage-based inverse problem for images can be viewed as a multi-objective optimization problem which involves the following three criteria:

- $F_1(z) = z^T A z + b^T z + c,$
- $F_2(z) = \sum_{i=1}^{N} H(z_i)$, where $H(z_i) = 1$ if $z_i > 0$ and 0 otherwise,
- $F_3(z) = \sum_{i=1}^{N} \frac{z_i}{Z} \ln \left(\frac{z_i}{Z} \right)$ where $Z = \max\{z_i\}$.

The function F_1 is the squared collage distance and it has to be minimized. . As mentioned above, the function F_2 measures the lack of sparsity of z and therefore has to be minimized. Finally the function F_3 is the negative Shannon entropy, which will have to be minimized. (Here we mention that other definitions of entropy could, in principle, be used.)

We are now interested in the solution to the following multi-objective problem,

$$\min_{z \in \Omega}(F_1(z), F_2(z), F_3(z)),\qquad(13)$$

where Ω is a compact subset of set $\mathbb{R}^{2N}$ which describes the set of all possible constraints. Because of presence of F_2 this optimization problem is nonsmooth and NP hard. We then proceed by replacing the function F_2 with a differentiable approximation. For this purpose, we propose the following approximation of $H_i(z)$,

$$H_\alpha(z_i) := 1 - \exp(-\alpha z_i^2), \quad \alpha > 0,\qquad(14)$$

and the accuracy of this approximation increases as $\alpha \to +\infty$.

There are different techniques to deal with a multi-objective optimization problem and the notion of optimal solution has to be understood in the Pareto sense. In fact it is unlikely to be able to determine an optimal vector z_{min} which minimizes all criteria simultaneously. However in practical situations, the simplest and most common approach which is used in this context is the one based on a scalarization technique. The idea behind this approach consists of combining all different criteria in a unique objective function by introducing a vector of non-negative weights, $L = (l_1, l_2, l_3)$, $\sum_{i=1}^{3} l_i = 1$. The scalarization of the above multi-objective model leads to the following single-criterion optimization problem

$$\min_{z \in \Omega} l_1 F_1(z) + l_2 F_2(z) + l_3 F_3(z).\qquad(15)$$

NUMERICAL SIMULATIONS

The following examples illustrate that adding small-weighted entropy and sparsity constraints can lead to a better fixed point approximation.

Example 1. For $j = 1, \dots, N$, we introduce the family of IFS maps

$$\begin{cases} w_i^j(x) &= \frac{1}{2^i}x + (j-1)\frac{1}{2^i}, \\ \phi_i^j(t) &= \alpha_i^j t + \beta_i^j, \end{cases} \quad i = 1, \dots, 2^j,$$

which for each j defines an associated IFSM map

$$(T_j u)(x) = \sum_{i=1}^{2^j}{}' \phi_i^j(u((w_i^j)^{-1}(x))).$$

The map T_j assembles 2^j shrunken and adjusted copies of $u(x)$, each supported on an interval of width $\frac{1}{2^j}$, of

the function u. For fixed j, the domains of the maps w_i^j only overlap at the endpoints of their domains. On the other hand, any point $x \in [0,1]$ that is not a multiple of $\frac{1}{2^m}$ for some m appears in the domain of exactly N of the maps in the family, once per member of the family, offering a sort of map redundancy. We define the combined (contractive) map

$$(Tu)(x) = \sum_{j=1}^{N}(T_j u)(x),$$

and consider the associated squared collage distance $F_1(z)$, where z is the vector of parameters α_i^j and β_i^j. In this example, we explore the scalarized optimization problem (15).

We choose the target function $v(x) = 0.8x^2 + 0.1$, $x \in [0,1]$. We set $N = 4$, which means we have a total of 60 parameters in the optimization problem. We use the nonlinear program solver in Maplesoft's Maple to solve (15). The first row of Table 1 shows the results without any entropy or sparsity constraints; we use 38 nonzero parameters. In the Table, values are presented with numbers of decimal places that illustrate the effect of changing the weights l_1, l_2, and l_3, and $\|v - \bar{u}\|_2$ is the L^2 distance between the target $v(x)$ and the resulting fixed point approximation $\bar{u} = T\bar{u}$.

Table 1. *Results of Example 1.*

l_2	$l_3 \,(\times 10^{-5})$	$\sqrt{F_1(z)}$
0	0	0.014300
0	1.3	0.014409
0	1.4	0.014420
0	1.5	0.014435
0.3	0	0.756211
0.01	0	0.066448
0.0001	0	0.016935
0.0001	1.5	0.016358

l_2	$F_2(z)$	$F_3(z)$	$\|v - \bar{u}\|_2$
0	38	-7.808589	0.0000741868
0	60	-12.648761	0.0000739184
0	60	-12.560957	0.0000731701
0	60	-12.701199	0.0000730789
0.3	6	-2.348598	0.0029801527
0.01	8	-2.545824	0.0003091702
0.0001	10	-2.984062	0.0000789527
0.0001	23	-7.929161	0.0000754027

In the absence of a sparsity constraint, *i.e.*, when $l_2 = 0$, we see that adding an entropy constraint with relatively small weighting l_3 leads to parameter values corresponding to a larger collage distance but to a better fixed point approximation. The numerical
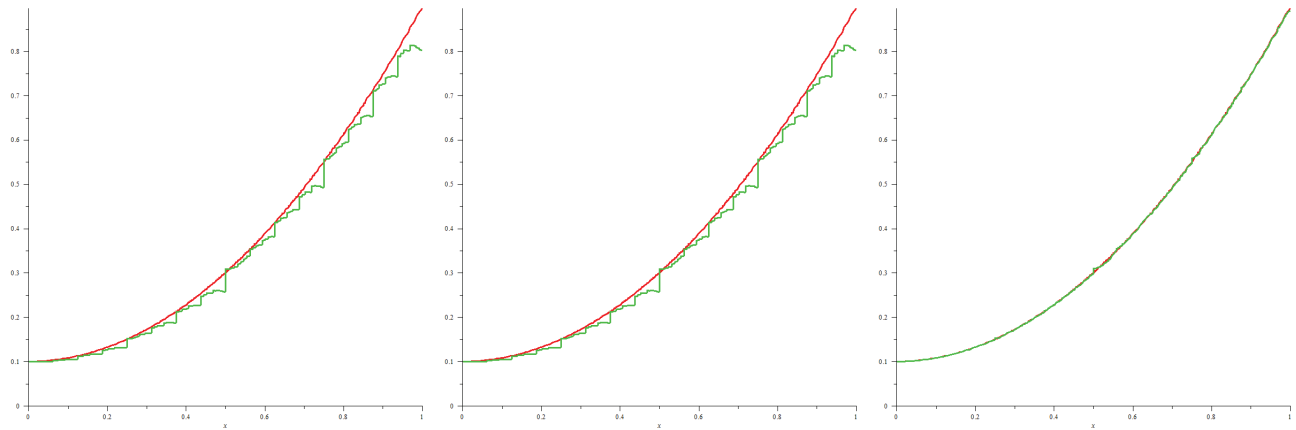
Fig. 1. *(left-to-right) Iterations 2, 4, and 8 of 23-map IFSM operator corresponding to the final row of Table 1.*

results are presented in rows two through four of the table; note that all 60 parameters are nonzero in these cases. Including only sparsity constraints, as in rows five to seven of the table, we see that the number of nonzero parameters is reduced dramatically from the collage distance case. Row five presents results for a case where only six parameters are nonzero, but the approximation error suffers. In row seven, we demonstrate that we can come close to the approximation error in row one while only using 10 nonzero parameters. Finally, in row eight, we include both entropy and sparsity constraints and come very close to the approximation error in row one by using less than two thirds of the number of parameters. Fig. 1 displays some iterates of the IFSM operator T corresponding to the final row of the table.

Example 2. We repeat the process of Example 1, this time using a nonmonotone target function $v(x) = \sin \pi x$ and setting $N = 6$ (corresponding to 124 parameters). Table 2 presents the results.

Table 2. *Results of Example 2.*

l_2	l_3 ($\times 10^{-7}$)	$\sqrt{F_1(z)}$
0	0	0.286458
0	5	0.286458
0	1	0.286458
0.0001	1	0.286500

l_2	$F_2(z)$	$F_3(z)$	$\|v - \bar{u}\|_2$
0	62	-19.707141	0.001142806379840
0	65	-23.653037	0.001142806381003
0	65	-23.651937	0.001142806379732
0.0001	35	-9.814576	0.001142959788729

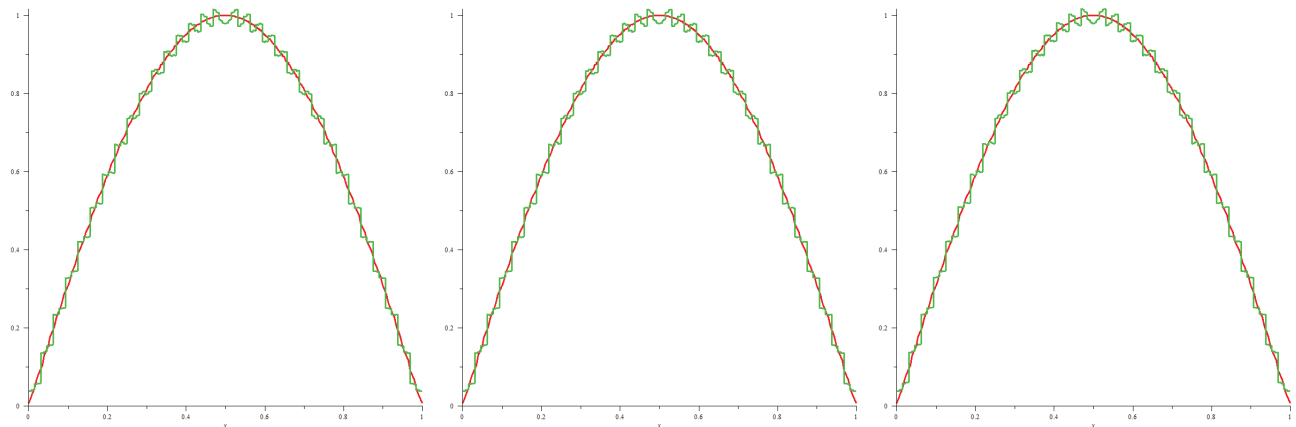In Fig. 2, we show some iterates of the IFSM operator T corresponding to the final row of the table.

Fig. 2. *(left-to-right) Iterations 2, 4, and 8 of 23-map IFSM operator corresponding to the final row of Table 2.*

REFERENCES

Barnsley MF, Ervin V, Hardin D, Lancaster J (1985). Solution of an inverse problem for fractals and other sets. Proc Nat Acad Sci USA 83:1975–77.

Barnsley MF (1989). Fractals everywhere. New York: Academic Press.

Barnsley MF, Demko S (1985). Iterated function systems and the global construction of fractals. Proc Roy Soc London Ser A 399:243–75.

Barnsley MF, Hurd L (1993). Fractal image compression. Natick, Mass: AK Peters.

Centore P, Vrscay ER (1994). Continuity of attractors and invariant measures for Iterated Function Systems. Canad Math Bull 37(3):315–29.

Demers M, Kunze H, La Torre D (2012). On random iterated function systems with greyscale maps. Image Anal Stereol 31(2):109–20.

Elad M. (2010) Sparse and redundant representations: From theory to applications in signal and image processing. New York: Springer.

Fisher Y (1995). Fractal image compression, theory and application. New York: Springer.

Forte B, Vrscay ER (1995). Solving the inverse problem for function and image approximation using iterated function systems. Dyn Contin Discret Impuls Syst 1(2):177-232.

Forte B, Vrscay ER (1998). Theory of generalized fractal transforms. In: Fisher Y, ed. Fractal Image Encoding and Analysis. NATO ASI Series F 159:145–68. New York: Springer Verlag.

Ghazel M, Freeman GH and Vrscay ER (2003). Fractal image denoising. IEEE Trans Image Proc 12(12):1560–78.

Hutchinson J (1981). Fractals and self-similarity. Indiana Univ J Math 30:713–47.

Iacus S, La Torre D (2005). A comparative simulation study on the IFS distribution function estimator. Nonlinear Anal Real World Appl 6(5):858–73.

Iacus S, La Torre D (2005). Approximating distribution functions by iterated function systems. J Appl Math Dec Sci 1:33–46.

Jaynes ET (1957). Information theory and statistical mechanics. Phys Rev 106(4):620–30.

Kunze H, La Torre D, Vrscay ER (2008). From iterated function systems to iterated multifunction systems. Commun Appl Nonlinear Anal 15(4):1–15.

Kunze H, La Torre D, Vrscay ER (2009). A generalized collage method based upon the Lax-Milgram functional for solving boundary value inverse problems. Nonlinear Anal 71(12):e1337–43.

Kunze H, La Torre D, Vrscay ER (2012). Solving inverse problems for DEs using the collage theorem and entropy maximization. Appl Math Letters 25(12):2306–11.

Kunze H, La Torre D, Mendivil F, Vrscay ER (2012). Generalized fractal transforms and self-similar objects in cone metric spaces. Comput Math Appl 64:1761–9.

La Torre D, Vrscay ER, Ebrahimi M, Barnsley M (2009). Measure-valued images, associated fractal transforms and the affine self-similarity of images. SIAM J Imaging Sci 2(2):470–507.

La Torre D, Vrscay ER (2009). A generalized fractal transform for measure-valued images. Nonlinear Anal 71(12):e1598–607.

La Torre D, Vrscay ER (2011). Generalized fractal transforms and self-similarity: recent results and applications. Image Anal Stereol 30(2):63–76.

La Torre D, Vrscay ER (2012). Fractal-based measure approximation with entropy maximization and sparsity constraints. AIP Conf Proc 1443:63–71.

Lu N (2003). Fractal imaging. New York: Academic Press.

Vrscay ER, Saupe D (1999). Can one break the 'collage barrier' in fractal image coding? In: Dekking M, LevyVehel J, Lutton E, Tricot C eds. Fractals: Theory and Applications in Engineering, London: Springer-Verlag. 307–23.