ABSTRACT

Image reconstruction from ray projections is a common technique in medical imaging. In particular, the few-view scenario, in which the number of projections is very limited, is important for cases where the patient is vulnerable to potentially damaging radiation. This paper considers swarm-based reconstruction where individuals, or particles, swarm in image space in an attempt to lower the reconstruction error. We compare several swarm algorithms with standard algebraic reconstruction techniques and filtered back-projection for five standard test phantoms viewed under reduced projections. We find that although swarm algorithms do not produce solutions with lower reconstruction errors, they generally find more accurate reconstructions; that is, swarm techniques furnish reconstructions that are more similar to the original phantom. A function profiling method suggests that the ability of the swarm to optimise these high dimensional problems can be attributed to a broad funnel leading to complex structure close to the optima. This finding is further exploited by optimising the parameters of the best performing swarm technique, and the results are compared against three unconstrained and boxed local search methods. The tomographic reconstruction-optimised swarm technique is shown to be superior to prominent algebraic reconstructions and local search algorithms.

CCS CONCEPTS
• Computing methodologies → Continuous space search; • Theory of computation → Bio-inspired optimization.

KEYWORDS
tomographic reconstruction, swarm optimisation, function profiling

1 INTRODUCTION

Tomographic Reconstruction (TR) is the inference of the internal structure of an object from the projected images cast by penetrating radiation [23]. TR plays an essential computational role in all medical imaging procedures (X-Ray CT, PET, MRI, Nuclear Medicine and ultrasound [20]). It also has wide applications in industry and science (data compression and data security [27], image processing [42], electron microscopy [11], crystal structure [7], angiography [12], nondestructive testing of homogeneous objects [22], seismic tomography [38], astronomy [9]) and geometric, combinatorial and recreational mathematics [6, 18].

The number of projections acquired in the imaging process is usually insufficient for a unique reconstruction. The problem is under-determined; furthermore, the random nature of the radiation, the operating characteristics of the detectors and, in medical applications, patient movement, mean that projection data is incomplete and noisy. The development of fast and accurate reconstruction algorithms that can identify structures from procedures with reduced patient radiation dose (by weakening the incident radiation and/or reducing acquisition time) is an important concern, enabling procedures that were once prohibited, and accelerating the throughput of patients. Diagnostic investigation would become possible in cases where it might have been withheld. These include imaging of young children, and where a patient would otherwise be required (for the requirements of a particularly clear image) to remain completely still, and to not even breathe for several minutes.

Various numerical reconstruction procedures have been developed. The Filtered Backprojection (FBP) algorithm is capable of fast reconstruction in a single iteration but requires a large number of projections [20]. Algebraic Reconstruction Techniques (ART) [27] are iterative algorithms based on Kaczmarz’s method [42] for solving an underdetermined system of equations. ART demonstrably reduces image noise and holds the potential for few-view reconstructions in a medical context but is subject to overfitting and evidence of their worth in large patient populations is lacking [19]. Exact reconstruction is possible if the data is transformed to a sparse and therefore compressible representation. Compressed Sensing (CS) exploits this principle and shows promise for few-view reconstruction, but the technique is dependant on knowledge of the sparse representation, and on replacing the non-convex optimisation problem typically unsolvable by traditional methods, by a solvable convex minimisation [10].

Solution techniques based on iterative statistical methods are also popular. The idea is to maximise the likelihood of parameters of an underlying statistical model (e.g. maximum likelihood expectation maximisation algorithm or MLEM [15]). MLEM converges slowly – an accelerated form known as ordered subset expectation maximisation (OSEM) [26] is faster – and images tend to become noisy. Noise is a symptom of overfitting and is a common problem with iterative techniques. Regularisation methods, such as maximum a posteriori (MAP) estimation, are often used to mitigate noise. Deep learning (DL) has revolutionised visual and natural language processing in recent years and there is a possibility that...
algorithms can be transferred and adapted for TR; however a radical improvement over analytical methods for solving inverse problems in imaging is so far missing [35, 46].

Conventional methods (e.g. FBP, ART, OSEM, CS and DL) have the potential for improved imaging (clearer images and reconstructions from reduced projections) but this potential has yet to be realised.

TR can be reformulated as a (typically) non-convex optimisation task: find a solution that minimises the error between the reconstructed image and the measured projections. Population-based, metaheuristic algorithms are often employed in optimisation and especially in cases where exact methods fail. One such metaheuristic is exemplified by particle swarm optimisation (PSO) [40]. Particles are simple agents that inhabit the space of feasible solutions; they have a memory of the best, as measured by the objective function, previously visited position and are subject to forces that drive them towards their personal best the best attained positions of social neighbours. PSO, by virtue of its simplicity, malleability, and wide applicability, has been the subject of much experimentation since its inception in 1995. It has been applied to 26 different problem categories [40] and around 100 papers are listed in a recent report on the application of computational swarm intelligence to real-world problems [2].

The solution set of the typical reconstruction problem is not unique and, although many reconstructions have small or zero error, only a few might correspond to medically useful images. ART and other least squares techniques will converge on a minimum norm (the sum of the squares of each pixel value) solution. Solutions with many small values will therefore dominate solutions with high values and the reconstructed image may be diffuse and lack structure. Swarm algorithms make no assumptions about the nature of the solutions. They do not require any special condition such as sparsity and are not driven towards a least squares solution.

There has been some investigation of heuristic TR algorithms. Ouaddah [39] combined harmony and local search. Other work (e.g. [13, 28, 29]) report on tabu search, simulated annealing and a memetic reconstruction, respectively. Batenberg [6] considered an evolutionary framework. Several authors report on swarm intelligent TR, covering areas such as: binary reconstruction [36], geophysical reconstruction [43], electrical capacitance and impedance tomography [25, 47] and surface reconstruction from 3D data [16]. Additionally PART, based on the movement of particles, was proposed for binary reconstruction [4].

This paper reports on a comparative study of swarm optimised TR, classic ART and FBP, for five standard benchmarks in the medical domain. We find that although the two population metaheuristics trialled (swarm optimisers and differential evolution) generally find solutions with higher reconstruction error than the best classical technique, they exhibit lower reproduction error i.e. the found solutions are closer to the original image. The observation that metaheuristics are unencumbered by the need to find a least squares solution might play a part in this finding; otherwise, a profiling study suggests that the objective function has a single broad funnel at large scales leading to complex multimodality close to the global optimal. This observation is further investigated by tailoring the swarm optimiser parameters to the TR context. The optimised swarm technique outperforms a number of local search methods and maintains its leading achievements against the best performing algebraic reconstruction approaches.

It is shown the tomographic reconstruction-optimised swarm technique demonstrates a lead in performance against the leading algebraic reconstruction approaches and the local search algorithms.

2 TOMOGRAPHY AND ALGEBRAIC RECONSTRUCTION

The imaging process proceeds by shining radiation on an object and collecting emergent radiation with a bank of detectors. The radiation source is swung around in a series of projection angles and emergent levels are recorded for each projection. The task is to infer interior structures (which absorb more or less radiation) from the projected images. The physical situation is exactly modelled by Radon transforms but transform inversions are infeasible. Instead, an approximate model (the ’forward’ model) of the physical measurement must be built in order to formalise the mathematical reconstruction problem.

2.1 Problem statement

Incident beams are typically modelled by parallel rays. Each ray is incident on the centre of each detector or projection bin. The imaging process is approximated by a projection matrix \( A \in \mathbb{R}^{m \times n} \) where \( m \) is the total number of rays collected (equal to the number of rays at each projection angle multiplied by the number of projection angles) and \( n \) is the number of pixels in the reconstructed image. If \( b \in \mathbb{R}^m \) is a vector of detector values, the continuous/discrete reconstruction problem can be stated as:

\[
\text{find } x \left( \in \mathbb{R}^n \right) \in \{0, 1, \ldots, k - 1\}^n, k > 1 \text{ such that } Ax = b. \quad (1)
\]

The binary problem is \( k = 2 \) i.e. with \( x \in \{0, 1\}^n \).

Since the equation \( Ax = b \) is, in general, underdetermined, it cannot be inverted. Instead an approximate solution \( y \) must be obtained (e.g. by filtered back projection, or by algebraic reconstruction). This trial solution is forward projected according to the measurement model:

\[
Ay = c
\]

with an associated reconstruction error

\[
e_1(y) = ||b - c||_1
\]

An iterative scheme will produce a sequence of candidate solutions, \( y^{(k)}, k = 1, 2, \ldots \) of non-decreasing error.

A zero projection error might yield a reconstructed solution \( y \) that is not identical to the original object \( x^* \). This is due to underdetermination. However, in cases where the reference image is known, the proximity of \( y \) to \( x^* \) offers a second measure of algorithm performance:

\[
e_2 = ||y - x^*||_1
\]
2.2 Algebraic reconstruction algorithms

The classical back projection [24, 32] technique, although a relatively quick and effective reconstruction procedure, suffers from high frequency ‘blurring’ which is only partly ameliorated by filters (FBP). However, increasing computation power means that algebraic reconstruction techniques (algebraic-RT or ART) are gaining prominence. This is due to ART’s potential for greater accuracy, albeit at increased time of execution.

The first ART was a rediscovery [21] of the Kaczmarz method for solving linear equations [31]; a diagrammatic explanation of the principle is provided in [33]. ART reconstructions suffer from salt and pepper noise, an artefact partially due to successive updates to components of \( x \) during an iteration changing the results of previously tuned values. This pathology is mitigated by the more slowly converging simultaneous iterative reconstruction technique (SIRT), where the updates are not applied immediately but are averaged and applied at the end of an entire iteration. In a further development, SART [5] improves upon ART and SIRT with the employment of a more sophisticated forward model. Good quality reconstructions are often obtained in a single iteration [33]. SART remains popular to this day and has been the subject of mathematical analysis (e.g. [30]).

3 SWARM OPTIMISATION

The essence of an optimisation swarm for continuous problems is a population of individuals or particles that move in a real-valued \( D \)-dimensional search space \( X \subset \mathbb{R}^D \). Particles respond to each other positions and value as determined by the objective function \( f : X \rightarrow \mathbb{R} \). The swarm moves through \( X \), typically clustering around promising areas before converging on a putative solution to the problem: find \( \min \ f(x) \).

The precise way that a particle alters its position in response to its neighbours, and any internal structure that it might possess, depends on the particular variety of swarm. For example, particles might carry a memory of a previously visited location; they might possess a velocity and they might interact with other particles in a pre-defined network, or with a random selection of neighbours chosen at each iteration.

The following outlines three swarm varieties: canonical particle swarm optimisation (PSO), a PSO-variant known as DFO and differential evolution (DE). Although DE is generally considered to be an evolutionary algorithm, it fits into the broad view of an optimisation swarm as outlined above.

3.1 PSO

Particles \( i \) in a canonical PSO swarm [34, 35] of \( N \) particles have dynamical variables \( x_i, v_i \), representing position and velocity in the search space \( X \subset \mathbb{R}^D \), and an internal ‘memory’ or pbest (personal best), \( p_i \), of the best position achieved so far in the run, as measured by the objective function \( f \).

Dynamical variables are updated according to the rule

\[
\begin{align*}
    v_i^{t+1} &= w v_i^t + c_1 (p_i^t - x_i^t) + c_2 (p^t - x_i^t) \\
    x_i^{t+1} &= x_i^t + v_i^{t+1}
\end{align*}
\]

where \( u_{1,2} \sim U(0,1) \) are uniform random variables in \([0,1]^D\) and \( \circ \) is the Hadamard (entry-wise) product. \( n_i \) is the pbest of the best neighbour in \( i \)’s social network, the inertial weight, \( w \), and acceleration coefficients \( c_1, c_2 \), are two arbitrary (but constrained) positive real parameters chosen to balance convergence and exploration and \( t \) labels iteration. In synchronous updating, iteration \( t+1 \) begins by updating all pbests:

\[
p_i^{t+1} = \min^* (f(x_i^t), f(p_i^t))
\]

where \( \min^* \) returns the first member of the list in the case of non-uniqueness. The iteration is completed by updating all \( N \) positions and velocities according to Eq. 3.

Two communication schemes are in common use. Particles in the global-best PSO (GPSO) network have access to all memories: the social network is maximally connected; on the other hand, particles in a local-best PSO (LPSO) network can only access memories in a restricted network. Networks do not include self. In the commonly chosen ring LPSO, particles communicate with ‘right’ and ‘left’ neighbours. LPSO, by virtue of a slower information transfer that inhibits convergence and favours early exploration, is generally better at more complex multi-modal problems [8].

3.2 DE

Differential evolution exists in a wide variety of forms; we specify the DE/best/1 version which is considered competitive and robust [14].

Each iteration begins with a determination of the current position, \( g \), of the best particle. Then, for each particle \( i \), indices \( j \) and \( k \) are selected such that \( i \neq j \neq k \). A random component \( r \in \{1, 2, \ldots, D\} \) is also selected.

Then for each component \( d \) of \( x_i \):

\[
\begin{align*}
    y_d &= g_d + F(x_{jd} - x_{kd}) \\
    y_d &= x_{id}
\end{align*}
\]

where \( y \) is a trial position and the parameters \( C_R \in [0,1] \) and \( F \in [0,2] \) are known as the ‘cross-over rate’ and the ‘differential weight’.

Then, after each component of \( y \) has been set, \( i \) is conditionally moved:

\[
x^{t+1} = \min^* (f(y), f(x_i^t))
\]

3.3 DFO

DFO, ‘dispersive flies optimisation’ [1], is a slimmed-down PSO variant that abolishes particle memory and velocity in favour of updates based on instantaneous, rather than historical, position; the algorithm’s exploration and exploitation behaviour is studied in [3]. In addition, it incorporates component-wise particle jumps [7]. The iteration starts by determining the best overall position \( g^{t+1} \), if unique, and positions of all best ring neighbours, \( n_i^{t+1} \) of each particle (except for the current swarm best particle, which is not updated). An arbitrary choice of \( g \) is made if there is a tie for the best position. Position component \( d \) of all particles \( i \), (other than
The potential of swarm reconstruction was tested with three swarm algorithms, PSO (in two varieties, GPSO and ring LPSO), DE and DFO; finally, since the swarm algorithms rely on extensive sampling, random search (RS) was also tested as a control.

The swarm algorithms and RS were run for 100,000 function evaluations. ART, CGLS, FBP, SART and SIRT perform the reconstruction in $\mathbb{R}^D$, where $D = 32 \times 32$ or $64 \times 64$. Reconstructions were scaled to $X = [0,255]^D$ for the purpose of computing the reproduction error, $e_2$.

A swarm size of $N = 100$ was chosen for G/LPSO, DE and DFO. Particles were initialised in $X$ with the uniform distribution and G/LPSO velocities were set to zero. Particles in all three swarms were clamped to the search box: any particle attempting to leave $X$ was placed on the boundary.

The DFO jump probability was set to 0.001; G/LPSO was run with $w = 0.729844$ and $c = 1.49618$ and the DE/best/1 parameters $F$ and $C_r$ were both set to 0.5.

All algorithms with randomisation were run 30 times on each of the 40 problems (5 phantoms, 4 projection types (6, 8, 16, 32) and 2 sizes, $32 \times 32$ and $64 \times 64$).

### 4 EXPERIMENTS AND RESULTS

The phantoms are depicted in Fig. 1, where phantoms 1 - 4 are binary reconstruction problems and phantom 5, the Shepp-Logan phantom, is a discrete problem with six pixel value levels [41]. Two sizes, $32 \times 32$ and $64 \times 64$, were trialled and, in order to test few-view conditions, the number of projections, $\alpha$, was set to 6, 8, 16 and 32. Phantom imaging was conducted by the ASTRA toolbox [45] using parallel geometry with the number of rays set to 32 and 64 for the $32 \times 32$ and $64 \times 64$ phantoms respectively.

Five reference algorithms from the ASTRA toolbox were selected: filtered backprojection (FBP), the algebraic algorithms ART, SIRT and SART, and a gradient descent reconstruction procedure, CGLS. The potential of swarm reconstruction was tested with three swarm algorithms, PSO (in two varieties, GPSO and ring LPSO), DE and DFO. Finally, since the swarm algorithms rely on extensive sampling, random search (RS) was also tested as a control.

Table 1 reports on Wilcoxon statistical significance tests on the reconstruction error for algorithm pairs for the 40 problem instances at a significance level of 0.05. The rows show the number of instances in which the row algorithm performed better than the column algorithm. For example, reading along the first row, ART gave a significantly smaller reconstruction error, $e_1$, than SART on 9 of the 40 trials.

Algebraic reconstruction algorithms ART, SART and SIRT and gradient descent, CGLS, uniformly outperform DE, GPSO and LPSO and are better than DFO in 32 trials. The result is not surprising;
Table 1: Algorithms comparison based on $e_1$. The numbers indicate statistically significant wins for the algorithm in the left hand column versus the algorithm in the top row.

<table>
<thead>
<tr>
<th>$e_1$</th>
<th>ART</th>
<th>CGLS</th>
<th>FBP</th>
<th>SART</th>
<th>SIRT</th>
<th>DE</th>
<th>DFO</th>
<th>GPSO</th>
<th>LPSO</th>
<th>RS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ART</td>
<td>NA</td>
<td>0</td>
<td>40</td>
<td>9</td>
<td>0</td>
<td>40</td>
<td>32</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>CGLS</td>
<td>40</td>
<td>NA</td>
<td>40</td>
<td>30</td>
<td>21</td>
<td>40</td>
<td>32</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>FBP</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>16</td>
<td>9</td>
<td>38</td>
<td>13</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>SART</td>
<td>31</td>
<td>10</td>
<td>40</td>
<td>NA</td>
<td>0</td>
<td>40</td>
<td>32</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>SIRT</td>
<td>40</td>
<td>19</td>
<td>40</td>
<td>NA</td>
<td>32</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>DE</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>40</td>
<td>4</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>DFO</td>
<td>6</td>
<td>6</td>
<td>30</td>
<td>6</td>
<td>40</td>
<td>NA</td>
<td>32</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>GPSO</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>LPSO</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>0</td>
<td>28</td>
<td>8</td>
<td>40</td>
<td>NA</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>RS</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>40</td>
<td>40</td>
</tr>
</tbody>
</table>

Table 2: Algorithms comparison based on $e_2$. The numbers indicate statistically significant wins for the algorithm in the left hand column versus the algorithm in the top row.

<table>
<thead>
<tr>
<th>$e_2$</th>
<th>ART</th>
<th>CGLS</th>
<th>FBP</th>
<th>SART</th>
<th>SIRT</th>
<th>DE</th>
<th>DFO</th>
<th>GPSO</th>
<th>LPSO</th>
<th>RS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ART</td>
<td>NA</td>
<td>31</td>
<td>37</td>
<td>29</td>
<td>28</td>
<td>5</td>
<td>3</td>
<td>38</td>
<td>3</td>
<td>40</td>
</tr>
<tr>
<td>CGLS</td>
<td>9</td>
<td>NA</td>
<td>22</td>
<td>15</td>
<td>13</td>
<td>7</td>
<td>3</td>
<td>25</td>
<td>6</td>
<td>37</td>
</tr>
<tr>
<td>FBP</td>
<td>3</td>
<td>18</td>
<td>NA</td>
<td>16</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>32</td>
<td>1</td>
<td>40</td>
</tr>
<tr>
<td>SART</td>
<td>11</td>
<td>25</td>
<td>NA</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>29</td>
<td>0</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>SIRT</td>
<td>12</td>
<td>27</td>
<td>36</td>
<td>22</td>
<td>NA</td>
<td>5</td>
<td>1</td>
<td>38</td>
<td>3</td>
<td>40</td>
</tr>
<tr>
<td>DE</td>
<td>34</td>
<td>31</td>
<td>39</td>
<td>40</td>
<td>35</td>
<td>NA</td>
<td>0</td>
<td>40</td>
<td>8</td>
<td>40</td>
</tr>
<tr>
<td>DFO</td>
<td>37</td>
<td>36</td>
<td>40</td>
<td>40</td>
<td>38</td>
<td>40</td>
<td>NA</td>
<td>32</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>GPSO</td>
<td>2</td>
<td>13</td>
<td>4</td>
<td>11</td>
<td>2</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>LPSO</td>
<td>36</td>
<td>34</td>
<td>39</td>
<td>40</td>
<td>36</td>
<td>28</td>
<td>8</td>
<td>NA</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>RS</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>40</td>
<td>40</td>
</tr>
</tbody>
</table>

these toolbox algorithms have been specifically developed for reconstruction whereas the swarm algorithms are off-the-shelf multi-purpose low-dimensional optimisers that have not been tuned to the reconstruction task.

The performance of swarm algorithms LPSO and DE is comparable to the classic filtered backprojection method; GPSO is notably worse and DFO has a slight advantage.

All algorithms are better than random searches which shows that the performance of the swarm algorithms is not merely due to repeated sampling.

Table 2 compares algorithm pairs when rated according to the reproduction error, $e_2$. DE, LPSO and DFO consistently find better reconstructions than any of the toolbox algorithms, and of the swarm algorithms, DFO exhibits a better performance. GPSO infrequently finds a lower $e_2$. DE and LPSO, although producing worse reconstructions when judged by the projection error, $e_1$, than the toolbox algorithms, nevertheless produce images that are closer to the original. However, the visual comparison in Fig. 1 shows that low reproduction error is not always a reliable measure on how a clinician might interpret an image; the toolbox algorithms produce more recognisable, albeit blurred, reconstructions of phantom 4 and the Shepp Logan phantom. DFO produce sharper images of phantoms 1-3. The blurriness seen in the toolbox reconstructions seems to have been traded for noise.

Table 3 presents a more detailed account of the $e_2$ results. The cells display the median reproduction error in the set of 40 trials, for each phantom; lighter shading indicates lower $e_2$. DFO produces good reproductions of phantoms 1-3, especially for higher numbers of projections: indeed it achieves perfect reproduction in seven cases. CGLS tends to have a lower median than the other toolbox algorithms, specially with higher number of projections, and SIRT is more consistent than SART. However, no toolbox algorithm is capable of producing an exact reconstruction of the original phantom.

Fig. 2 shows $e_1$ convergence plots for sample runs. The plots do not show evidence of stagnation, a characteristic morbidity of swarm algorithms, and indicate that extended runs would have reduced the errors still further.

4.2 Function profiles

A surprising feature of the experiments is the ability of the swarm algorithms, as exemplified by the convergence plots of Fig. 2, to make any progress given the high dimensionality and perceived difficulty of the reconstruction problems.

It is worth noting that the swarm algorithms of these trials have been developed for comparably low-dimensional problems (typically in 30 dimensions) whereas the reconstructions here, at 1024 and 4096 dimensions, are considered high in the global optimisation community and are subject to special methods.

In order to test the ability of a general swarm optimiser i.e. one that has not been enhanced for high dimensional problems, trial runs of DFO on a blank phantom (in which all pixel values are zero) and on the unimodal Sphere problem in 1024 dimensions were performed. Fig. 5, a representative convergence plot, shows the swarm does indeed make progress and does not suffer any periods of stagnation. The shape of the plots differs from the phantom convergence plots of Fig. 2 in which the initial fast convergence is followed by a slowing down (the plots are convex rather than concave), hinting that the problem is unimodal at large scales but has a more complex multimodal structure on small scales.

The nature of the task was investigated further by measuring function profiles along ‘adaptive’ walks. An adaptive walk is a sequence of steps in $X$ such that each step lowers or equals the function value of its predecessor. We define an adaptive walk profile of granularity $\lambda$ as a plot of objective function values taken at $1/\lambda$ points in a straight line between steps in an adaptive walk.

Fig. 3 shows adaptive walks profiles at iterations $t = 200$ to $t = 210$. The adaptive walk is the trace of the swarm best position at each iteration. The ‘anchor steps’ (actual values attained by the swarm) are marked with black blobs. The continuous black line follows $e_1$ values at $1/\lambda = 100$ intermediate points. In all five cases we see that the profile never rises above the value at the end-points, indicating that the function, as seen by the swarm, is unimodal at this scale, and particles do no encounter ridges between optima and the swarm is contained within a single funnel.

Fig. 4 shows profiling at later iterations. The objective functions appears to lose their apparent unimodality close to the optima at
Table 3: Reproduction error, $e_2$, for each problem and each algorithm. Lighter shading indicates the proximity of the reconstructions to the phantoms. The largest errors in phantoms of sizes $32^2$ and $64^2$ are $255 \times 32^2$ and $255 \times 64^2$ respectively.

<table>
<thead>
<tr>
<th>Phantom, size = $32^2$, $\alpha = 6$</th>
<th>ART</th>
<th>CGLS</th>
<th>FBP</th>
<th>SART</th>
<th>SRT</th>
<th>DE</th>
<th>DFO</th>
<th>GPSO</th>
<th>LPSO</th>
<th>RS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 8$</td>
<td>51836</td>
<td>103862</td>
<td>82141</td>
<td>52578</td>
<td>52254</td>
<td>24630</td>
<td>8925</td>
<td>90896</td>
<td>17006</td>
<td>132172</td>
</tr>
<tr>
<td>$\alpha = 16$</td>
<td>51730</td>
<td>53772</td>
<td>87373</td>
<td>52574</td>
<td>53452</td>
<td>18306</td>
<td>1825</td>
<td>88231</td>
<td>12444</td>
<td>122473</td>
</tr>
<tr>
<td>$\alpha = 32$</td>
<td>38732</td>
<td>59710</td>
<td>55565</td>
<td>56237</td>
<td>42305</td>
<td>10800</td>
<td>0</td>
<td>87165</td>
<td>8898</td>
<td>125350</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phantom, size = $64^2$, $\alpha = 6$</th>
<th>ART</th>
<th>CGLS</th>
<th>FBP</th>
<th>SART</th>
<th>SRT</th>
<th>DE</th>
<th>DFO</th>
<th>GPSO</th>
<th>LPSO</th>
<th>RS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 8$</td>
<td>125181</td>
<td>146809</td>
<td>426382</td>
<td>214874</td>
<td>246654</td>
<td>219940</td>
<td>196317</td>
<td>435764</td>
<td>154123</td>
<td>506633</td>
</tr>
<tr>
<td>$\alpha = 16$</td>
<td>227777</td>
<td>352616</td>
<td>386348</td>
<td>225147</td>
<td>241197</td>
<td>193099</td>
<td>179152</td>
<td>432352</td>
<td>137348</td>
<td>507789</td>
</tr>
<tr>
<td>$\alpha = 32$</td>
<td>163363</td>
<td>473576</td>
<td>263650</td>
<td>193599</td>
<td>197971</td>
<td>217275</td>
<td>161105</td>
<td>430594</td>
<td>117964</td>
<td>507951</td>
</tr>
</tbody>
</table>

$e_1 = 0$. The profile shows a complex structure of ridges where intermediate points exceed the values at the end-points and the profiles are very jagged. At this stage in the optimisation, the swarm has to negotiate micro-basins that separate sub-optimal minima.

The emergent picture of $e_1$ as a single broad funnel leading to a highly modal surface close to the global optimum provides some explanation of the relative success of any swarm algorithm to make progress in this high dimensionality, and of superior performance of LPSO over GPSO, given LPSO is known to have better multimodal characteristics [8]. DFO, which is uniformly better than DE and GPSO (see Table 1) is also more successful than DE on these reconstruction problems. Given the (formal) intermediate nature of DFO between PSO and DE, it seems that the removal of velocity and history, and the retention of a mixed (local and global) communication network is a superior strategy.

4.3 Parameter fine-tuning

The parameters of the leading swarm method were optimised in order to better gauge its performance in the context of the problems. Phantom 1, with size $32 \times 32$ and $\alpha = 6$, was used as a benchmark for the sweep through parameter space. The three DFO parameters were selected from the sets $N = \{2, 3, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}, \phi = \{0.17, 0.346, 0.520, 0.693, 0.866, 1.039, 1.212, 1.386, 1.559, 1.732\}$ and $\Delta = \{0.0, 0.00001, 0.00005, 0.0001, 0.0005, 0.001, 0.01, 0.05, 0.1\}$. The search was first applied to find the optimum $N$ and $\phi$ setting. Each parameter combination was run 10 times. The best values found were $N = \{2, 3\}$ and $\phi = \{1.386, 1.559, \sqrt{3}\}$, with $N = 2$ having a small lead (i.e. performing statistically significantly better in [118,117,117] cases for each $\phi$ values out of the total of 130 algorithms, as opposed to $N = 3$, with [117,111,113] cases). Subsequently, using $N = 2$, and $\phi = 1.386$ the search for the optimum restart thresholds resulted in an equal performance for $\Delta = \{0.0005, 0.001\}$. 

a) al-Rifaie and Blackwell
Figure 2: Reconstruction error, $e_1$, during 1,000 iterations in sample runs for the phantoms.

Figure 3: Unimodal function profiles

Figure 4: The revelation of complex function profiles at later iterations

Figure 5: DFO error in 1000 iterations in a sample run for a blank phantom ($32 \times 32$) where all pixels are set to zero, and the Sphere function, $e = \sum_{d=1}^{D} X_d^2$, with $X$ constrained to $[0, 255]^D$ and $D = 1024$.

Grid-based search for G/LPSO with $N \in \{2, 3, \ldots, 100\}$ and DE with $N \in \{3, 4, \ldots, 100\}$, resulted in $N = 100$ in both cases which indicates that small populations are not favoured by these algorithms.

To have a more precise picture of the optimal parameter values for $N$, $\phi$, and $\Delta$, DFO was used as a hyper-heuristic. The experiment was run 30 times for the 3 dimensional problem, with the population size of 10 and the termination criterion set to 100 iterations (1000 function evaluations). In this experiment, the elitism mechanism was relaxed to allow for the re-evaluation of the current best parameter set in each iteration, followed by a best individual update if necessary. The median values of the parameters found in the trials are then extracted. The optimum values found were $N = 2$, $\phi = 1.7320508 \approx \sqrt{3}$ and $\Delta = 0.0011245 \approx 0.001$, in agreement with the sweet spot in the grid search.

While the small value of optimal $N$ might come as a surprise, it confirms the function profiling which demonstrated the presence of a single broad funnel leading to small scale complex structure close to global optima, in effect rendering the task into a largely
unimodal problem. It also indicates that although DFO is acting as a ‘swarm-inspired local search’ and the collective presence of a large communication network is unnecessary in this context.

4.4 Local search and algorithm comparison

To further explore this finding, the tomographic reconstruction-optimised DFO (DFO-TR), with $N = 2, \phi = \sqrt{3}$ and $\Delta = 0.001$, is compared against three local search algorithms (Nelder-Mead [37], L-BFGS-B [48], MTS-LS1 [44]) and the best performing toolbox algorithm, SIRT. Nelder-Mead is unconstrained while L-BFGS-B and MTS-LS1 are boxed in the feasible space. MTS-LS1 is especially designed for large scale global optimisation, and the classic L-BFGS-B uses an approximation of the gradient to improve the search. MTS-LS1 is appropriated for separable problems, but is sensitive to rotations. On the other hand, L-BFGS-B is less powerful but is less sensitive to rotations. Nelder-Mead is trialled in two flavours, the Standard Nelder-Mead Simplex (SNMS) [37] and the Adaptive Nelder-Mead Simplex (ANMS) [17]. ANMS has been proposed to deal with the inefficiency of SNMS in high dimensions.

All algorithms were run 30 times on each of the 40 problems, with each trial running for 100,000 function evaluations.

Table 4 compares these algorithms based on the reconstruction ($e_1$), and reproduction ($e_2$) errors. L-BFGS-B and Nelder-Mead algorithms show no promise in the context of these high dimensional problems. SIRT is the best algorithm in terms of the reconstruction error as shown in Table 4-top, followed by DFO-TR, and then MTS-LS1. However, in terms of the reproduction error (see Table 4-bottom), DFO-TR is the leading method, followed by MST-LS1 and then SIRT. The low reproduction error of DFO-TR, in comparison to the other algorithms (toolbox and local search techniques), places this algorithm in a promising spot for dealing with reconstruction problems. More work is required to establish if there are factors other than clamping, restart and local search which contribute to the performance of the investigated methods.

![Figure 6: Phantom size: 32 × 32, projections: 6](image)

5 CONCLUSION

This paper reports on swarm optimised tomographic reconstruction of five standard phantoms in the few-view regime. Three representative swarm algorithms were tested and compared with algebraic reconstruction algorithms and filtered backprojection, along with two unconstrained and two boxed local search algorithms. We find that DFO – an algorithm that formally interpolates between the widely studied and applied differential evolution and particle swarm optimisation techniques – is the most competitive swarm method. The standard toolbox techniques provide lower reconstruction error and blurred final images, whereas swarm algorithms produce lower reproduction error (sometimes even producing exact reconstructions) and sharper images peppered with noise.

An adaptive walk function profiling technique suggests that these reconstruction problems consist of a single broad funnel leading to small scale complex structure close to global optima. This large-scale unimodality would enable swarms to make rapid progress towards the optima. Convergence plots confirm that the swarms do not suffer from stagnation.

Following the profiling and evident large-scale unimodality, the TR-optimised DFO or DFO-TR is compared against a number of local search methods. The results demonstrate the better performance of this small-swarm algorithm against both the local search methods and the best performing toolbox algorithm. More detailed studies are required to establish the reasons behind the outcome which could lie in the use of clamping, the restart mechanism, and the overwhelming occupancy of the optima on the edges.

Furthermore, a hybridisation of classical toolbox techniques (with their low reconstruction error) and swarm optimisation (returning low reproduction error) could produce a powerful algorithm capable of fast reconstructions in the few view regime.

REFERENCES
