

Gradient-Based Swarm Optimization for ICA



Rasmikanta Pati, Vikas Kumar and Arun K. Pujari

Abstract Blind source separation (BSS) is one of the most interesting research problems in signal processing. There are different methods for BSS such as principal component analysis (PCA), independent component analysis (ICA), and singular value decomposition (SVD). ICA is a generative model of determining a linear transformation of the observed random vector to another vector in which the transformed components are statistically independent. Computationally, ICA is formulated as an optimization problem of contrast function, and different algorithms for ICA differ among themselves on the way the contrast function is modeled. Several optimization techniques such as gradient descent and variants, fixed-point iterative methods are employed to optimize the contrast function which is nonlinear, and hence, determining global optimizing point is most often impractical. In this paper, we propose a novel gradient-based particle swarm optimization (PSO) method for ICA in which the gradient information along with the traditional velocity in swarm search is combined to optimize the contrast function. We show empirically that, in this process, we achieve better BSS. The paper focuses on the extraction of one by one source signal like deflation process.

Keywords ICA · Contrast function · Optimization · Gradient Particle swarm optimization

R. Pati (✉)
SUIIT, Sambalpur University, Sambalpur, India
e-mail: rkpati@suiit.ac.in

V. Kumar · A. K. Pujari
School of CIS, University of Hyderabad, Hyderabad, India
e-mail: vikas007bca@gmail.com

A. K. Pujari
e-mail: akpujari@curaj.ac.in

A. K. Pujari
Central University of Rajasthan, Kishangar, Ajmer 305817, Rajasthan, India

1 Introduction

The independent component analysis (ICA) is one of the most prominent methods of data analysis and has been widely used in signal processing, pattern recognition, and machine learning. In signal processing, ICA is essentially viewed as a computational method for separating multivariate signals into additive components and has been applied in many contexts. ICA is extensively used in pattern recognition and image analysis mainly in applications like face recognition, object recognition, image filtering, embedding in feature space. ICA is similar to PCA in many respect, but unlike PCA, ICA attempts to determine which are independent. The inherent advantage of ICA is its ability to recover source (or unobserved signal) from observed mixture. ICA is also popularly known as a method of blind source separation (BSS). It is called blind because we do not have information about the source signals or the mixing method. For BSS, it is assumed that signals from source can be mixed linearly or nonlinearly. ICA attempts to separate source by some simple assumptions of their statistical properties. In ICA, data are represented by the random vector x and the components as the random vector s . It is to determine a transformation that maps observed data x into maximally independent components s for some measure of independence. The transformation is usually assumed to be linear, and the measure of independence can be a measure of non-Gaussianity.

Let us consider an observed M -dimensional discrete time signal where the n th sample is denoted by the column vector $x(n)$. The observed signal is the mixture of unknown N -dimensional source vector $s(n)$ given by

$$x(n) = As(n) \tag{1}$$

where A is a linear mixing matrix. The input components are usually statistically dependent due to the mixing process, whereas the sources are not. If one succeeds in finding a matrix W that yields statistically independent output components $y(n)$, given by

$$y(n) = Wx(n) = WAs(n) \tag{2}$$

one can recover the original sources up to a permutation and constant scaling of the sources. W is called the demixing matrix, and finding the matrix is referred to as independent component analysis (ICA).

ICA computation involves determining the matrix W by a process of optimization of a non-convex optimization, and the widely adopted gradient descent algorithms [1] usually converge to a local optimizing point and seldom find the global optimizing point. As no global solution is guaranteed, most of ICA techniques exhibit random behavior yielding different results for different initial conditions and initial values of parameters. There are different approaches of estimating W . Maximization algorithm based on singular value decomposition (SVD) [2], gradient optimization of kurtosis function [3], and iterative method [5] of approximating W . In this paper, we are particularly examining the deflation-based source separation.

In this paper, a particle swarm optimization (PSO)-based ICA algorithm is presented to overcome the above problem. As an evolutionary computation technique and general global optimization tool, PSO was first proposed by Kennedy and Eberhart [4] which simulates the simplified social life models. Since PSO has many advantages over other heuristic techniques such as it can be easily implemented and has a great capability of escaping local optimal solutions [5], PSO has been applied successfully in many computer science and engineering problems. Another drawback of gradient-based methods [9, 14] is slow speed of convergence. PSO search is preferred over gradient search when the nonlinear objective function is multimodal and there are a large number of local optimizing solutions. In such a situation, gradient-based search gets stuck at a local optimizing point where as population based technique search through a broader area ensuring t possibility of reaching global optimizing solution. So an obvious question is whether one can combine gradient information of search direction together with the velocity computed by local/global best solution to enhance the search. Taking advantages of both gradient search and population based search, we propose a method which blends gradient search with PSO. We show empirically that by this process, we can have an efficient method of ICA computation.

The rest of the paper is organized as follows. In Sect. 2, we briefly review the cumulants, reference signal, contrast function. Section 3 discusses the optimization technique and iterative procedure for ICA. A brief introduction about particle swarm optimization (PSO) is given in Sect. 4. Section 5 describes our proposed methods termed as PSOAS for ICA. Experimental analysis of the proposed method is reported in Sect. 6. Finally, Sect. 7 concludes and indicates several issues for future work.

2 Cumulant-Based Contrast Optimization

A contrast function is any nonlinear function which is invariant to permutation and scaling matrices and attains its minimum value in correspondence of the mutual independence among the output components. Many contrast functions for ICA has been proposed in the literature, mainly based on information theoretical principles such as maximum likelihood, mutual information, marginal entropy, and negentropy, as well as related non-Gaussianity measures. Among them, the kurtosis (normalized fourth-order marginal cumulant) is arguably the most common statistics used in ICA, even if skewness has also been proposed.

Statistical properties of the output dataset can be described by its moments or, more conveniently, by its cumulants. Since the data have zero mean, the sample cumulants up to order four can be written in the following way.

$$\begin{aligned}
 C_i^{(y)} &= 0 \\
 C_{ij}^{(y)} &= \langle y_i y_j \rangle \\
 C_{ijk}^{(y)} &= \langle y_i y_j y_k \rangle \\
 C_{ijkl}^{(y)} &= \langle y_i y_j y_k y_l \rangle - \langle y_i y_j \rangle \langle y_k y_l \rangle - \langle y_i y_k \rangle \langle y_j y_l \rangle - \langle y_i y_l \rangle \langle y_j y_k \rangle
 \end{aligned}$$

with $\langle \cdot \rangle$ indicating the mean over all data points.

Cumulants of a given order form a tensor. The diagonal elements characterize the distribution of single component and the fourth-order autocumulant, $C_{iiii}^{(y)}$ is kurtosis of y_i . The cross-cumulants characterize the statistical dependencies between components. Thus, if and only if, all components are statistically independent, the off-diagonal elements (or the cross-cumulants) vanish. ICA is equivalent to finding an unmixing matrix W that diagonalizes the cumulant tensors of the output data at least approximately. Though it is easy and trivial to achieve diagonalization of first- and second-order cumulants, there is no obvious way of diagonalizing higher-order cumulant tensors. The diagonalization of these tensors can only be done approximately, and we need to define an optimization criterion for this approximate diagonalization

The approximate diagonalization of the cumulant tensors of order three and order four is achieved by minimizing an objective function which is the sum of the squared third- and fourth-order off-diagonal elements. Since the sum of square of all elements of a cumulant tensor is preserved under any orthogonal transformation of the underlying data, one can equivalently maximize the sum over the diagonal elements instead of minimizing the sum over the off-diagonal elements. This is a contrast function as defined in [6]. Thus, the process can be viewed as an optimization problem with the following objective function.

$$J(y) = \frac{1}{\mu} \sum_{\alpha} (C_{\alpha\alpha\alpha}^{(y)})^2 + \frac{1}{\tau} \sum_{\alpha} (C_{\alpha\alpha\alpha\alpha}^{(y)})^2, \tag{3}$$

The objective function J is kurtosis [7, 8] and can be rewritten as a function of an orthogonal matrix U which is to be determined through the optimization process. Expressing the above criterion function in terms of U is not straightforward, and hence, another cumulant-based contrast function is defined as follows. This definition uses cumulant of order four only. Recently, *reference-based* contrast functions are proposed based on cross-statistics or cross-cumulants between the estimated outputs and reference signals. Reference signals are nothing but artificially introduced signals for facilitating the maximization of the contrast function. Due to the indirect involvement of reference signals in the iterative optimization process, these reference-based contrast functions have an appealing feature in common: The corresponding optimization algorithms are quadratic with respect to the searched parameters.

$$\begin{aligned}
 C_z\{y\} &\triangleq \text{Cum}\{y, y, z, z\} \\
 &= E\{y^2 z^2\} - E\{y^2\}E\{z^2\} - 2E^2\{yz\}
 \end{aligned} \tag{4}$$

where $E\{\cdot\}$, denotes the expectation value and z is the reference signal. We consider another(reference) separation matrix V and $z(n) = Vx(n)$. We now define the contrast function explicitly in terms of W and V as follows.

$$I(W, V) = \left| \frac{C_z\{y\}}{E\{(y)^2\}E\{(z)^2\}} \right|^2 \quad (5)$$

where $y(n) = Wx(n)$ and $z(n) = Vx(n)$.

3 Optimization Method

There have been a umpteen number of proposals to optimize the contrast function defined previously. In order to avoid an exhaustive search in the whole space of orthogonal matrices, a gradient ascent on $J(U)$ is normally used. The gradient of a function is the vector of its partial derivatives. It gives a direction of the maximum increase in the function leading to an update rule for U that looks like

$$U(k+1) \leftarrow U(k) + \lambda(k)\nabla J|_{U(k)} \quad (6)$$

where $\nabla J|_{U(k)}$ denotes either the natural, or relative gradient of J with respect to U , evaluated at $U = U(k)$.

Gradient ascent and its variants start with a random seed point and move from one point to another in the gradient direction. The performance of all gradient-based approach depends on a factor such as step size λ and initial seed point $U(0)$. The rate of convergence highly depends on the selection of step size, and an improper step size may lead to the poor performance and stability of the algorithm.

Use of a gradient-based maximization supposes that the algorithm will not be trapped in a spurious maximum, leading to U^* , that does not correspond to a satisfactory solution for the BSS problem (still mixing). Various authors such as [9, 10] have noted that the usual ICA contrast functions may have such spurious maxima if several source distributions are multimodal. For instance, Cardoso in [11] shows this phenomenon for the likelihood-based contrast function. More recently, Vrins et al. [12] have given an intuitive justification regarding the existence of spurious maxima when the opposite of the output marginal entropies is used for the contrast function.

A simple gradient search algorithm for the maximization of kurtosis-based contrast J is given in Algorithm 1. It is shown in [3] that Algorithm 1 may diverge unacceptably leading to a numerical overflow if a great number of iterations are required by the algorithm.

Algorithm 1:

input : $x(n)$: Observed signal
output: v : Separation vector

Initialize randomly U_0

for $k = 0, 1, \dots, k_{max} - 1$ **do**

$d_k = \nabla I(U_k, U_k)$

$\alpha_k = \underset{\alpha}{\operatorname{argmax}} I(U_k + \alpha d_k, U_k)$

$U_{k+1} \leftarrow U_k + \alpha_k d_k$

end

The separating property is not affected by a scaling factor, because of unavoidable scaling ambiguity in BSS [3]. It is common in BSS to impose the unit-power constraint $E\{|y(n)|^2\} = 1$. It is known that the unit-power constraint is equivalent to a unit-norm constrain on the separating vector v . A modified algorithm to avoid the drawback of Algorithm 1 is proposed in [3] by normalization of the separating vector v after every gradient iteration update. The points found after renormalizing the above algorithm belong to unit sphere. The main flow of the modified algorithm can be found in Algorithm 2. We use this method in our comparative studies in the later section.

The output obtained after the maximization process should be closer to the source signal rather than the reference one. With this aim, a modification is proposed in [9] where the reference vector is updated after each iteration by the output signal computed in the previous iteration.

Algorithm 2:

input : $x(n)$: Observed signal
output: v : Separation vector

Initialize U_0 .

for $k = 0, 1, \dots, k_{max} - 1$ **do**

$d_k = \nabla_1 I(U_k, U_k)$

$\alpha_k = \underset{\alpha}{\operatorname{argmax}} I(U_k + \alpha d_k, U_k)$

$\tilde{U}_{k+1} \leftarrow U_k + \alpha_k d_k$

$\hat{U}_{k+1} \leftarrow \frac{\tilde{U}_{k+1}}{(E\{|\tilde{U}_{k+1}\{x(n)\}^2\})^{\frac{1}{2}}}$

$U_{k+1} \leftarrow \hat{U}_{k+1}$

end

4 Particle Swarm Optimization

Particle swarm optimization (PSO) is a well-known population-based search. The PSO algorithm works by simultaneously maintaining several candidate solutions in the search space to find the global optimum, where the movement is influenced by the social component and the cognitive component of the particle. The most characteristic feature of PSO and its variants is that the search trajectory is influenced by the best solutions (local best and global best) obtained so far in the search to determine the next solution. Each individual particle has a velocity vector v_i , a position vector x_i , personal best pb_i that the particle encountered so far and neighborhood best lb_i means the best position that all particles have encountered so far among the neighborhood N_i of particle i . The position and velocity of each particle are updated as follows.

$$\begin{aligned} v_i^{t+1} &= v_i^t + c_1 r_1 (pb_i^t - x_i^t) + c_2 r_2 (lb_i^t - x_i^t) \\ x_i^{t+1} &= x_i^t + v_i^{t+1} \end{aligned}$$

where c_1, c_2 are acceleration coefficient and $r_1, r_2 \in [0, 1]$ are uniformly distributed random numbers.

PSO is particularly attractive for its ability to yield global optimizing point with the fast converging rate. However, it does not use the gradient information which is very crucial for optimization.

5 PSOAS: The Proposed Method

In this section, we discuss the method of blending swarm search with gradient-based optimization for ICA. Unlike PSO, in the proposed algorithms, the velocity component of the particle is updated in every iteration with gradient direction along with the social influence. The search direction of the particle is a combination of gradient direction and the direction of global best. There have been some earlier proposals which use PSO to solve the ICA problem [6, 13, 14]. In the literature, many variants of gradient-based PSO exist [15–17]. Some researchers has combined a gradient factor with search direction computed by personal best and global best whereas in [18] terminates gradient search is initiated after termination of PSO.

Algorithm 3 describe the detail procedure related to applicability of PSO on Algorithm 1.

Algorithm 3: Gradient-based PSO

input : $x(n)$: Observed signal, S : Swarm size and δ : Trade-off parameter
output: U^{best} : Separation vector

Initialize U_0 and the corresponding reference signal $z_0^p(n) = U_0^p x(n), \forall 1 \leq p \leq S$.

for $k = 0, 1, \dots, k_{max} - 1$ **do**
 $I_p = I(U_k^p, U_k^p), \forall 1 \leq p \leq S$
 $best = \underset{p}{argmax} I_p$
 $d_k^p = \nabla I(U_k^p, U_k^p)$
 $\alpha^p = \underset{\alpha}{argmax} I(U_k^p + \alpha d_k^p, U_k^p)$
 $\tilde{U}_{k+1}^p \leftarrow U_k^p + \alpha^p (\delta d_k^p + (1 - \delta)(U^{best} - U_k^p))$
 $\tilde{U}_{k+1}^p \leftarrow \frac{\tilde{U}_{k+1}^p}{(E\{|\tilde{U}_{k+1}^p x(n)|^2\})^{\frac{1}{2}}}$
 $U_{k+1}^p \leftarrow \tilde{U}_{k+1}^p$
end

We modify Algorithm 3 with iterative updates to get another alternative, Algorithm 4 as follows.

Algorithm 4: Gradient-based PSO with Fixed-point update

input : $x(n)$: Observed signal, S : Swarm size and δ : Trade-off parameter
output: U^{best} : Separation vector

Initialize U_0 and the corresponding reference signal $z_0^p(n) = U_0^p x(n), \forall 1 \leq p \leq S$.

for $k = 0, 1, \dots, k_{max} - 1$ **do**
 $\tilde{U}_0^p = U_k^p, \forall 1 \leq p \leq S$
 for $l = 0, 1, \dots, l_{max} - 1$ **do**
 $I_p = I(\tilde{U}_l^p, U_k^p), \forall 1 \leq p \leq S$
 $best = \underset{p}{argmax} I_p$
 $\tilde{d}_l^p = \nabla_1 I(\tilde{U}_l^p, U_k^p)$
 $\tilde{\alpha}^p = \underset{\alpha}{argmax} I(\tilde{U}_l^p + \alpha \tilde{d}_l^p, U_k^p)$
 $\tilde{U}_{l+1}^p \leftarrow \tilde{U}_l^p + \tilde{\alpha}^p (\delta \tilde{d}_l^p + (1 - \delta)(U^{best} - \tilde{U}_l^p))$
 $\tilde{U}_{l+1}^p \leftarrow \frac{\tilde{U}_{l+1}^p}{(E\{|\tilde{U}_{l+1}^p x(n)|^2\})^{\frac{1}{2}}}$
 end
 $U_{k+1}^p \leftarrow \tilde{U}_{l_{max}}^p$
end

6 Simulation

This section discusses the experimental setup and reports the results. We conducted experiments on a variety of synthetic datasets. Complex-valued, independent, and identically distributed (i.i.d) QAM4 has been generated taking their values in

Table 1 Experimental results of each comparing algorithm in terms of average and median MSE

	Parameters		Number of samples							
			500		1000		5000		10000	
	k_{max}	l_{max}	PSOAS	GAS	PSOAS	GAS	PSOAS	GA	PSOAS	GAS
Average MSE	1000	1	0.0066	0.0114	0.0007	0.0047	0.0003	0.0039	0.0003	0.0043
	200	5	0.0005	0.0046	0.0316	0.0214	0.0007	0.0038	0.0007	0.0049
	100	10	0.0161	0.0098	0.0006	0.0049	0.0008	0.0044	0.0005	0.0034
	50	20	0.0010	0.0050	0.0236	0.0301	0.0005	0.0049	0.0007	0.0048
	25	40	0.0171	0.0109	0.0006	0.0065	0.0004	0.0035	0.0005	0.0047
	10	100	0.0010	0.0035	0.0342	0.0407	0.0004	0.0045	0.0005	0.0040
	8	125	0.0271	0.0099	0.0005	0.0061	0.0007	0.0051	0.0004	0.0043
	5	200	0.0007	0.0038	0.2894	0.3538	0.0004	0.0053	0.0004	0.0045
Median MSE	1000	1	0.0001	0.0009	0.0001	0.0005	0.0000	0.0006	0.0000	0.0006
	200	5	0.0000	0.0008	0.0002	0.0010	0.0000	0.0006	0.0000	0.0007
	100	10	0.0002	0.0009	0.0001	0.0007	0.0000	0.0005	0.0000	0.0005
	50	20	0.0001	0.0007	0.0002	0.0012	0.0000	0.0006	0.0000	0.0006
	25	40	0.0002	0.0006	0.0001	0.0008	0.0000	0.0005	0.0000	0.0006
	10	100	0.0001	0.0004	0.0003	0.0018	0.0000	0.0006	0.0000	0.0006
	8	125	0.0002	0.0006	0.0001	0.0010	0.0000	0.0007	0.0000	0.0006
	5	200	0.0001	0.0006	0.1834	0.2570	0.0000	0.0006	0.0000	0.0006

$\{e^{i\pi/4}, e^{-i\pi/4}, e^{+i3\pi/4}, e^{-i3\pi/4}\}$ with equal probability $\frac{1}{4}$. For a different number of sample, a set of $N = 3$ mutually independent and temporally i.i.d source has been generated. They have been mixed by a QL finite impulse response (FIR) filter with randomly driven coefficients of length 3 and with $Q = 4$ sensors. The separating FIR separator has been searched with length $D = N(L - 1) = 6$.

To measure the performance of different algorithms, we have employed mean squared error (MSE) as an evaluation metric popularly used in blind source separation [3]. We report the equalization performance of the proposed method by taking the average and median values of MSE of 1000 trials. We compare our proposed methods with two well-known algorithms: Algorithm 1 and Algorithm 2 [3] with our Algorithm 3 and Algorithm 4, respectively.

Table 1 gives the comparative analysis of proposed method against state-of-the-art algorithms on different datasets. The best results among all comparing algorithm are highlighted in boldface. The row corresponding to value of $k_{max} = 1000$ and $l_{max} = 1$ reports the results provided by Algorithm 1 [3] and Algorithm 3 proposed in the present work. The remaining rows show the results provided by Algorithm 2 [3] and Algorithm 4 proposed in the present work. It can be seen from the Table 1 that the proposed method achieves better performance consistently than other comparing algorithms in terms of each evaluation metric. The following tables denote the algorithms, Algorithm 1 and Algorithm 2 of [3], as general algorithms (GAS) and our proposed algorithms, Algorithm 3 and Algorithm 4, as PSO algorithms (PSOAS).

7 Conclusion

In this paper, two new algorithms have been developed with PSO-based search for the purpose of maximizing the kurtosis contrast function. Particularly, Algorithm 4 allows two parameters to improve performance for practical purpose. The work also opens for future works with respect to genetic algorithm and source separation of complex-valued signals based on nonlinear autocorrelation. One can use genetic algorithm to see its practical purpose. As well as PSO may use in the scenario of nonlinear autocorrelation.

References

1. Castella, M., Moreau, E.: A new method for kurtosis maximization and source separation. In: 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), pp. 2670–2673. IEEE (2010)
2. Kawamoto, M., Kohno, K., Inouye, Y.: Eigenvector algorithms incorporated with reference systems for solving blind deconvolution of mimo-iir linear systems. *IEEE Signal Process. Lett.* **14**(12), 996–999 (2007)
3. Castella, M., Moreau, E.: New kurtosis optimization schemes for miso equalization. *IEEE Trans. Signal Process.* **60**(3), 1319–1330 (2012)
4. Kennedy, J., Eberhart, R.: Particle swarm optimization (ps). In: Proceedings of IEEE International Conference on Neural Networks, Perth, Australia, pp. 1942–1948 (1995)
5. Parsopoulos, K.E., Vrahatis, M.N.: Recent approaches to global optimization problems through particle swarm optimization. *Nat. Comput.* **1**(2–3), 235–306 (2002)
6. Igual, J., Ababneh, J., Llinares, R., Miro-Borras, J., Zarzoso, V.: Solving independent component analysis contrast functions with particle swarm optimization. *Artif. Neural Netw. ICANN 2010*, 519–524 (2010)
7. Simon, C., Loubaton, P., Jutten, C.: Separation of a class of convolutive mixtures: a contrast function approach. *Signal Process.* **81**(4), 883–887 (2001)
8. Tugnait, J.K.: Identification and deconvolution of multichannel linear non-gaussian processes using higher order statistics and inverse filter criteria. *IEEE Trans. Signal Process.* **45**(3), 658–672 (1997)
9. Castella, M., Rhioui, S., Moreau, E., Pesquet, J.C.: Quadratic higher order criteria for iterative blind separation of a mimo convolutive mixture of sources. *IEEE Trans. Signal Process.* **55**(1), 218–232 (2007)
10. Boscolo, R., Pan, H., Roychowdhury, V.P.: Independent component analysis based on nonparametric density estimation. *IEEE Trans. Neural Netw.* **15**(1), 55–65 (2004)
11. Haykin, S.S.: *Unsupervised Adaptive Filtering: Blind Source Separation*, vol. 1. Wiley-Interscience (2000)
12. Vrins, F., Archambeau, C., Verleysen, M.: Entropy minima and distribution structural modifications in blind separation of multimodal sources. In: AIP Conference Proceedings, vol. 735, pp. 589–596. AIP (2004)
13. Krusienski, D.J., Jenkins, W.K.: Nonparametric density estimation based independent component analysis via particle swarm optimization. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05), vol. 4, pp. iv–357. IEEE (2005)
14. Li, H., Li, Z., Li, H.: A blind source separation algorithm based on dynamic niching particle swarm optimization. In: MATEC Web of Conferences. Volume 61., EDP Sciences (2016)
15. Borowska, B., Nadolski, S.: Particle swarm optimization: the gradient correction. (2009)

16. Noel, M.M., Jannett, T.C.: Simulation of a new hybrid particle swarm optimization algorithm. In: Theory, System (ed.) 2004, pp. 150–153. IEEE, Proceedings of the Thirty-Sixth Southeastern Symposium on (2004)
17. Vesterstrom, J.S., Riget, J., Krink, T.: Division of labor in particle swarm optimisation. In: Evolutionary Computation, 2002. CEC'02. Proceedings of the 2002 Congress on. Volume 2., IEEE (2002) 1570–1575
18. Szabo, D.: A study of gradient based particle swarm optimisers. PhD thesis, Masters thesis, Faculty of Engineering, Built Environment and Information Technology University of Pretoria, Pretoria, South Africa (2010)