A Novel Accurate Source Number Estimation Method Based on GBSA-MDL Algorithm

Taha Bouras(✉), Di He, Fei Wen, Peilin Liu, and Wenxian Yu

Shanghai Key Laboratory of Navigation and Location-Based Services, Shanghai Jiao Tong University, Shanghai, People's Republic of China
tahabouras04@yahoo.com,
{dihe, wenfei, liupeilin, wxyu}@sjtu.edu.cn

Abstract. Several classical source number estimation methods have been proposed in the past based on information theoretic criteria such as minimum description length (MDL). However, in most known real applications there is a scenario in which the number of sensors goes to infinity at the same speed as the number of snapshots (general asymptotic case) which yields to a blind performance for the classical MDL and results in an inaccurate source number estimation. Accordingly, in this work, the Galaxy Based Search Algorithm (GBSA) is modified and applied with the MDL criteria in order to optimize and correct the detection of source number under such sample-starving case. Simulation results show that the proposed GBSA-MDL based method gives reliable results compared to several used source number estimation methods.

Keywords: Source number estimation methods
Minimum Description Length (MDL) · General asymptotic case
Optimization · Galaxy Based Search Algorithm (GBSA)

1 Introduction

In recent decades, direction-of-arrival (DOA) or angle-of-arrival (AOA) estimation of multiple sources using an array is a fundamental problem in modern signal processing and it has found wide applications in radar, sonar, communications, geophysics, tracking, localization [1–4] and indoor positioning such as the application of AOA-based positioning through the use of wireless local area network (WLAN) infrastructure [5]. However, compared with different DOA estimation methods, the subspace-based methods are the most commonly used due to their high resolution and simplicity of computational such as multiple signal classification (MUSIC) [6] and estimation of signal parameters via rotational invariance techniques (ESPRIT) [7]. Meanwhile, in order to separate between source subspace and noise subspace which are the compositions of the signal impinging the array elements on the receiver side, the subspace-based algorithms require the exact information of the source number which is usually unknown.

Several classical source number estimation methods have been proposed in the past which are based on information theoretic criteria like the Akaike information criterion (AIC) [8], Bayesian information criterion (BIC) [9], predictive description length

B. Li et al. (Eds.): ChinaCom 2017, LNICST 237, pp. 383–392, 2018.
https://doi.org/10.1007/978-3-319-78139-6_39

(PDL) [10], Gerschgorin disk estimator (GDE) [11], and Minimum Description Length (MDL) [12]. Compared to other source estimation methods like, the decomposition of eigenvalues, methods based on information criteria do not require any parameters in order to fitly estimating the source number from the blended signals which allow them to be easier and efficient methods. However, it has been shown in [13, 14] that the MDL criteria suffer from the very low SNR as well as the sampling rate.

In the other side, the interaction between intelligent software tools and nature yields to a something new called Meta-heuristic algorithms. They are simply algorithms who imitate the behavior of nature in order to find a precise solution to hardly optimized problems. These type of methods (also named as nontraditional optimization methods) are known with their robustness for solving complicated engineering problems such as Simulated Annealing (SA) [15], Genetic Algorithm (GA) [16], Invasive Weed Optimization (IWO) [17] and Particle Swarm Optimization (PSO) [18] - as framework based on swarm behavior. Newly, a novel optimization algorithm was invented called Galaxy Based Search Algorithm (GBSA) which mimics the spiral galaxies in the outer space and it has gained more attention in a wide variety optimization problems, such as the use of GBSA to optimize the Otsu's criterion for multilevel thresholding of gray-level images in [19], GBSA for principal components analysis in [20] and a method for tracking an object using modified GBSA (M-GBSA) has been used in [21].

After constructing a complex objective function, the main problem of the information-based methods is to find its correct extremum values based on information criteria aspect. Otherwise, in real applications, there is a case when the number of sensors will be larger or equal to the number of samples that is known as the asymptotic case and this affects the performance of the MDL and yields to a non-detection for the source number or inaccurate source enumeration [13]. According to that, in this work, GBSA is used to solve this non-uniform optimum problem in order to enhance the shortcoming performance of the MDL under the effects of such starving environments.

The following of the paper is organized as follows. In Sect. 2, we set the problem formulation and the classical minimum description length. In Sect. 3, the main idea and the procedure for the GBSA are introduced. The proposed GBSA-MDL is described in Sect. 4. In Sect. 5, we present the simulation results and in the last Sect. 6, we draw conclusions for our work.

2 Problem Formulation and Minimum Description Length

In DOA estimation, the typical signal model is;

$$X(t) = A(\theta) * S(t) + w(t) \tag{1}$$

Where, time is represented as t^{th} sample such that $t = 1, \ldots, n$. $X(t) = [X_1(t) \ldots X_m(t)] \in \mathbb{C}^{m \times 1}$ is the observed snapshot vector. $A(\theta) = [a(\theta_1)\, a(\theta_2) \ldots a(\theta_d)] \in \mathbb{C}^{m \times d}$ is the steering matrix, where $a(\theta)$ is called steering vector and $\theta = [\theta_1, \theta_2, \ldots, \theta_d]^T$ are the parameters of interest or true DOA's.

$S(t) = [S_1(t)\ldots\ldots S_d(t)]^T \in \mathbb{C}^{d\times 1}$ is the d source vector and $w(t) = [w_1(t)\ldots \ldots w_m(t)]^T \in \mathbb{C}^{m\times 1}$ is the noise vector. $(.)^T$ represents the transpose operation, *d* is the unknown source number, m is the number of antennas at the receiver side and n is the number of snapshots. A fateful problem here is to estimate *d* from *n* finite set of observations before an efficient source separation by using the MDL criterion function based on the concept of the shortest description length for data done by Wax [12]. So for presumed sources, the MDL criterion is given as:

$$\mathrm{MDL}(k) = n(m-k)\log\left(\frac{\frac{1}{m-k}\sum_{i=k+1}^{m} l_i}{\left(\prod_{i=k+1}^{m} l_i\right)^{1/m-k}}\right) + \frac{1}{2}k(2m-k)\log n \tag{2}$$

Where l_i is the i^{th} eigenvalue of the covariance R, with $i = 1,\ldots, m$, and $l_1 \geq l_d \geq l_{d+1} \geq \ldots \geq l_m$, such that $R = E\left[XX^T\right]$ and E[.] is the expectation function. As we have mentioned before, the classical MDL criterion function described in (2) can provide a reliable performance just in the case when n tends to infinity while m remains fixed. Otherwise, the used criteria will easily fail. As result, The main problem of our work is to use the GBSA as a rescue algorithm, more clearly after modifying the GBSA to be more suitable for our optimization problem in the general asymptotic case, it is then used to find the correct source number $\hat{d}$ and this can be performed by minimizing the objective (fitness) function (2) with respect to k in different situations ($m = n$, $m < n$, and $m > n$).

$$\hat{d} = \arg_{k=0,\ldots\bar{m}-1} \min \mathrm{MDL}(k) \tag{3}$$

Where $\bar{m}$ is the upper limit of the source number. Generally, is fixed to $\min(m, n) - 1$.

3 GBSA

3.1 Origin of GBSA

Spiral galaxies are a certain kind of galaxy, which consists of a flat disk with a bulging center and surrounding spiral arms [20]. Figure 1 displays a Photograph for a spiral galaxy. As the galactic disk angular speed of turnover differs with distance from the galactic center, a radial arm (such as a spoke) would speedily become curved to tightly curved just after slight galactic rotations. According to that, the GBSA algorithm uses a spiral chaotic move in order to imitate this spiral arm. After each iteration, it will be more narrowly curved which allows reaching precisely to the optimum solution.

Recently, GBSA is becoming a breakthrough optimization method because of its simplicity of implementation as well as ability to swiftly converge to a convenient solution. It does not require any gradient information of the function to be optimized and uses only primitive mathematical operators. Also, GBSA is well suited to solve the non-linear, non-convex, continuous, discrete, integer variable type problems.

Fig. 1. An example of spiral galaxy

3.2 GBSA Procedure

A. *Initialization*
B. *Spiral Chaotic Move*

The spiral chaotic move is the procedure of the GBSA used to find the global optima. It follows the procedures shown in Fig. 2. When the initial solution SG is set randomly, the spiral chaotic move searches globally around the given solution in order to find the area where the peak point may locate and this is known as exploration. The initial solution SG is updated to SGnext and this happens continuously every time the updated solution is more superior to the current one. Since the multimodal problems suffer from the high existence of the local optimum, the GBSA algorithm uses a chaotic variable in order to avoid falling into the same solution and thus keep the chance of the variation in the solutions.

C. *Local search*

After exploring the searching area globally, the local search procedure would be activated to exploit the promising area where may the peak of the fitness function exists following the steps shown in Fig. 3.

4 The Proposed GBSA-MDL Source Estimation Method

The proposed GBSA-MDL is shown in Fig. 4. A major breakthrough in the new proposed algorithm takes advantage from the presumed source number of the MDL which relies on the dimension of GBSA, as a result ending up with an efficient source enumerator in the general asymptotic case.

First of all is to verify whether the number of sensors goes to infinity at the same speed as the number of snapshots, m, n $\rightarrow \infty$ and m/n $\rightarrow$ c ϵ $(0, \infty)$, if the condition is not satisfied then the classical MDL is called in order to find the number of sources frugally. Hence, GBSA can be used to find the solution even that the above-mentioned condition is not satisfied, however using the classical MDL is more convenient to avoid time conception as well as computation complexity we can see that more clearly in the simulation section.

```
θ ← -π
    While rep<MaxRep
        Repeat for i =1 to L
SNexti ← Si+NextChaos() rcos(θi)
Endrepeat
If (f(Snext)≥ (f(S)) then
Flag ← True
GotoEndprocedure
Endif
          Repeat for i=1 to L
SNexti← Si-NextChaos() rcos(θi)
Endrepeat
          If (f(Snext)≥ (f(S)) then
Flag ← True
GotoEndprocedure
Endif
r ← r+Δr
          Repeat for i=1 to L
θi← θi+Δθ
Endrepeat
          Repeat for i=1 to L
If (θi> π) then
θi← - π
Endif
Endrepeat
rep ←rep+1
Endwhile
Endprocedure
```

Fig. 2. The pseudo-code of *"spiral chaotic move"* procedure.

```
Repeat for i=1 to L
α← 1
k← 0
While k <kMax
SLi ← Si-α.ΔS.NextChaos()
SUi ← Si+ α .ΔS.NextChaos()
    If f(SL)<f(S) and f(SU)<f(S) then
GotoEndprocedure
Endif
    If f(SU)>f(S) then
Si ← SUi
SLi← SUi
α← α + 0.01 x NextChaos()
k← k+1
Else
f(SL)>f(S) then
Si ← SLi
SUi← SLi
α← α + 0.01 x NextChaos()
k← k+1
Else
α← α + 0.05 x NextChaos()
k ← k+1
Endif
Endifwhile
SLi← Si
SUi← Si
Endrepeat
SNext← S
Endprocedure
```

Fig. 3. The pseudo-code of *"local search"* procedure.

In the opposite side, when the general asymptotic case occurs, the classical MDL criteria will suffer from the multidimensionality, nonlinearity and complexity problems ending up with a blind estimation or maybe not detection at all.

However, in such sample-starving case, the rescue of the modified GBSA algorithm involves, it takes the following steps as shown in Fig. 4 in order to search for the optimum solution which is at the same time the number of sources needed:

(1) Over the searching space, the generation of initial random solutions should be determined using the "Generate Initial Solution" component of the GBSA, the presumed source number will be set as the dim of the GBSA.
(2) Evaluate the fitness of all agents using Eq. (3) which is determined by taking the data from all the elements of the array for n number of snapshots.
(3) The solution that GBSA-MDL seeks is the locus of the agent corresponding to minimum fitness which in fact represents the source number estimation. So, after step two we apply the GBSA algorithm. The turn of "Spiral Chaotic Move" is activated in order to explore the study area and confines the space where the maximum solution might be included.

After the discovery of the likely area by the "Spiral Chaotic Move", the Flag is set to true as shown in Fig. 4, next, the local search procedure starts to find locally new optimum solution better than its two immediate neighbors until it converges to the appropriate source number.

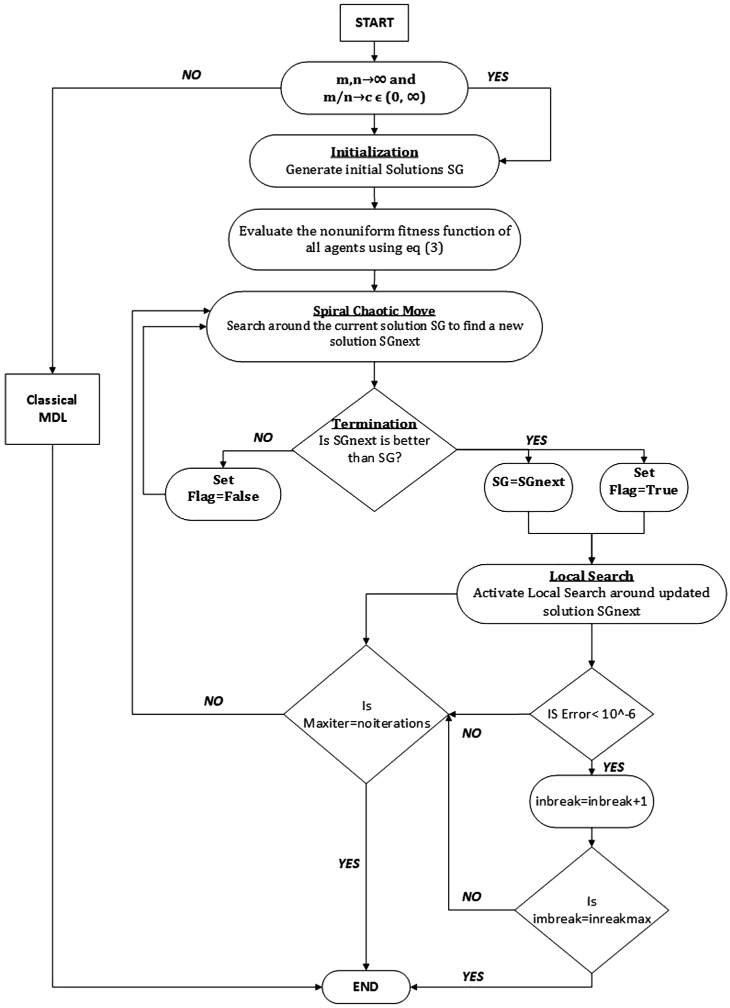

Fig. 4. Flowchart of the GBSA-MDL source estimation method

(4) The GBSA-MDL terminates when it reaches the following stopping criteria Fig. 4.
   (a) The GBSA-MDL stops when the iteration counting of the GBSA overruns Maxiter = 100.
   (b) The second stopping criterion examines whatever the difference error between two sequential solutions is less than the value of 1e−10. If it is true then inbreak = inbreak + 1. Hence, when ‘inbreak’ exceeds a predetermined value our algorithm halts computations.
   (c) The last stopping criterion is the number of trials and is set to be 100 in our simulations.

5 Simulation Results and Discussions

We consider a ULA of 30 sensors with half wavelength element separation impinged by three uncorrelated far-field narrow band Stationary Gaussian signals with angles [2.9°, 9.4°, 16.5°]. Figures 5, 6 and 7 show the performance of our proposed GBSA-MDL compared with some existing information criteria, called exponentially embedded families (EEF), RMT-AIC and corrected AIC (AIC_{c3}) in terms of probability of error detection versus SNR with ratio m/n = 1/3, 1, and 3 respectively such that m represents the number of antennas at the receiver side and n is the number of snapshots. Meantime, all numerical results are computed for 300 independent trials such that parameters of the GBSA used during the optimization process are given in Table 1.

When the sample size n is smaller than the number of sensors m Fig. 5, the proposed method outperforms the two schemes EEF and AIC_{c3}. It is obvious that the GBSA-MDL provides a consistent estimation for source number compared to EEF which cannot provide any detection. Also, the GBSA-MDL performs better than the AIC_{c3} in terms of probability of error detection. In addition of that, we can observe that our algorithm can proceed an accurate source detection while the RMT-AIC be unstable even when the SNR goes to infinity.

When the number of samples increases and is equal to m, all methods can provide a detection as shown in Fig. 6. The proposed GBSA-MDL, EEF and AIC_{c3} criteria can correctly detect the source number, GBSA-MDL still performs better than EEF, and almost proceed the same as the AIC_{c3}. In addition of that, EEF and AIC_{c3} need a higher SNR in order to end up with an accurate detection. Meanwhile, RMT-AIC gives unreliable estimation even that the number of SNR is larger compared to GBSA-MDL.

Table 1. GBSA parameters used for optimization.

Spiral chaotic move		Local search		GBSA	
Maxrep	50	***Kmax***	50	***Iteration number***	60
Drmax	1	***Ls_sstep***	0.0001	***Dimension***	6
		Ls_lstep	0.05		
Dtmax	0.1	***Dsmax***	10	***Chaosvar***	0.19

As illustrated in Fig. 7, when n becomes much larger than m the gap between the proposed GBSA-MDL and the two criteria EEF and AIC_{c3} is reliable which implies that the performance of GBSA-MDL is always consistent while RMT-AIC criterion is still suffering for reaching zero probability of detection even that at sufficiently SNR.

In this part, we want to prove the performance of the proposed GBSA-MDL in terms of snapshots compared to the classical MDL by considering a ULA with the same number of sensors m = 30 received two far-field narrow bands Stationary Gaussian signals from two different angles [0.5°, 3.5°], SNR is set to be −5 dB.

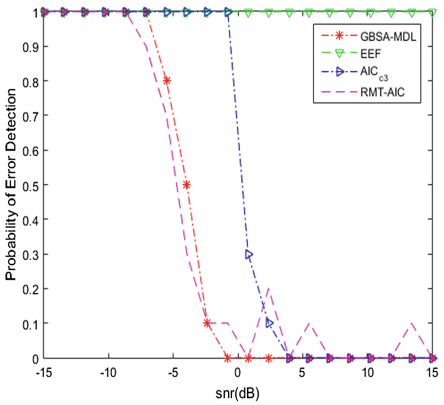

Fig. 5. Probability of error detection versus SNR with m = 30 and n = 10.

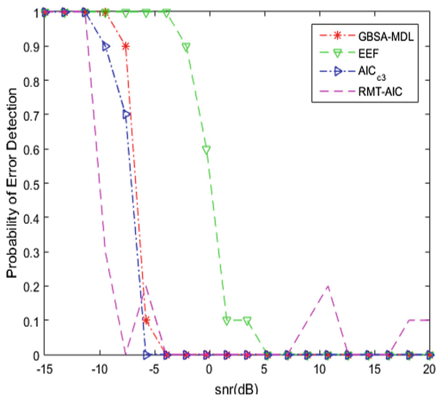

Fig. 6. Probability of error detection versus SNR with m = 30 and n = 30.

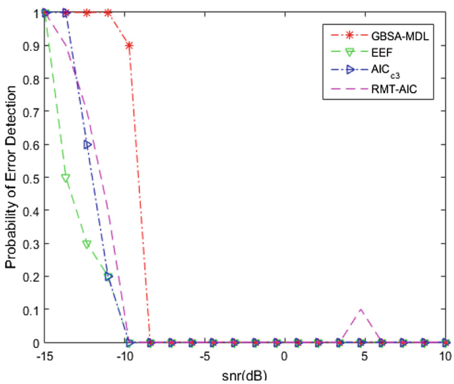

Fig. 7. Probability of error detection versus SNR with m = 30 and n = 90.

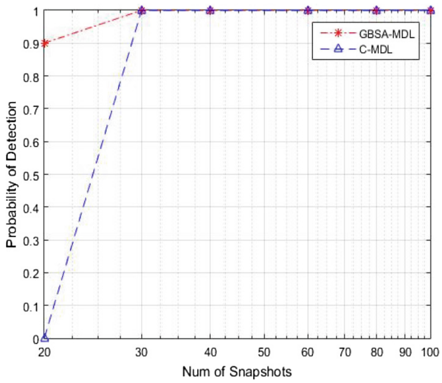

Fig. 8. Probability of detection versus number of snapshots. SNR = −5 db.

Figure 8 shows the probably of detection of source of interest with respect to the number of snapshots for the proposed GBSA-MDL and the Classical MDL. It is obvious from Fig. 8 that the GBSA-MDL outperforms the CMDL in terms of snapshots. When the number of snapshots is less than the number of sensors, $n < 30$, and the probability of correct detection for GBSA-MDL is almost one while CMDL suffers to reach the correct detection and this can improve the reliable detection of the GBSA-MDL in the general asymptotic case compared to the classical MDL. Meanwhile, the CMDL criterion needs around additional 30 snapshots in order to provide the same accuracy as GBSA-MDL.

6 Conclusion

In this work, a GBSA-MDL source number estimation method has been investigated by using the meta-heuristic GBSA Algorithm to improve the performance of the Classical MDL. Numerical results proved the usability of the GBSA-MDL for source enumerator in several environments. In addition to that, the comparison with the existing source enumerators (CMDL, EEF, RMT-AIC, AIC_{c3}) shows the capability and the robustness of the GBSA-MDL for providing an accurate estimation for source number in the general asymptotic case where the number of sensors tends to infinity at the same time as the number of snapshots whereas the other approaches fail in such starving environment.

Acknowledgment. This research work is supported by the Important National Science and Technology Specific Project of China under Grant No. 2016ZX03001022-006, the Shanghai Science and Technology Committee under Grant No. 16DZ1100402, and the National Natural Science Foundation of China under Grant No. 91438113.

References

1. Xie, W., Wen, F., Liu, J., Wan, Q.: Simultaneously source association, DOA and fading coefficients estimation for multipath signals. IEEE Trans. Signal Process. **65**(11), 2773–2786 (2017)
2. Wen, F., Wan, Q., Fan, R., Wei, H.: Improved MUSIC algorithm for multiple noncoherent subarrays. IEEE Signal Process. Lett. **21**(5), 527–530 (2014)
3. Wen, F., Xie, W., Chen, X., Liu, P.: DOA estimation for noncircular sources with multiple noncoherent subarrays. IEEE Commun. Lett. (2017). https://doi.org/10.1109/lcomm.2017.2703649
4. Xie, W., Wang, C., Wen, F., Liu, J., Wan, Q.: DOA and gain-phase errors estimation for noncircular sources with a central symmetric array. IEEE Sens. J. **17**(10), 3068–3078 (2017)
5. Wong, C., Klukas, R., Messier, G.: Using WLAN infrastructure for angle-of-arrival indoor user location. In: Proceedings of the 68th Semi-Annual IEEE Vehicular Technology (VTC 2008), pp. 1–5, September 2008
6. Schmidt, R.O.: Multiple emitter location and signal parameter estimation. IEEE Trans. Antenna Propag. **34**(3), 276–280 (1986)
7. Roy, R., Kailath, T.: ESPRIT-estimation of signal parameters via rotational invariance techniques. IEEE Trans. Signal Process. **37**(7), 984–995 (1989)
8. Akaike, H.: A new look at the statistical model identification. IEEE Trans. Autom. Control **AC-19**(6), 716–723 (1974)
9. Schwarz, G.: Estimating the dimension of a model. Ann. Stat. **6**(1978), 461–464 (1987)
10. Valaee, S., Kabal, P.: An information theoretic approach to source enumeration in array signal processing. IEEE Trans. Signal Process. **52**(5), 1171–1178 (2004)
11. Wu, H.T.: Source number estimators using transformed Gerschgorin radii. IEEE Trans. Signal Process. **43**(6), 1325–1333 (1995)
12. Wax, M., Kailath, T.: Detection of signals by information theoretic criteria. IEEE Trans. Acoust. Speech Signal Process. **33**(2), 387–392 (1985)

13. Huang, L., So, H.C.: Source enumeration via MDL criterion based on linear shrinkage estimation of noise subspace covariance matrix. IEEE Trans. Signal Process. **61**, 4806–4821 (2013)
14. Huang, L., Long, T., Mao, E., So, H.C.: MMSE-based MDL method for accurate source number estimation. IEEE Signal Process. Lett. **16**, 798–801 (2009)
15. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. Science **220**, 671–680 (1983)
16. Goldberg, D.E.: Genetic Algorithms in Search Optimization and Machine Learning. Addison-Wesley, Boston (1989)
17. Mehrabian, A.R., Lucas, C.: A novel numerical optimization algorithm inspired from weed colonization. Ecol. Inform. **1**, 355–366 (2006)
18. Kennedy, J., Eberhart, E.C.: Particle swarm optimization. In: IEEE International Conference on Neural Networks, vol. 4, pp. 1942–1948 (1995)
19. Shah-Hosseini, H.: Otsu's criterion-based multilevel thresholding by a nature-inspired metaheuristic called galaxy-based search algorithm. In: Third World Nature and Biologically Inspired Computing (NaBIC) (2011). 2013 MECS
20. Shah-Hosseini, H.: Principal components analysis by the galaxy-based search algorithm: a novel metaheuristic for continuous, optimisation. Int. J. Comput. Sci. Eng. **6**(1/2), 132–140 (2011)
21. Sardari, F., Moghaddam, M.E.: An object tracking using modified galaxy based search algorithm. Swarm Evol. Comput. **30**, 27–38 (2016)