

## Review Article

# A Perspective of Conventional and Bio-inspired Optimization Techniques in Maximum Likelihood Parameter Estimation

Yongzhong Lu<sup>1</sup>, Min Zhou<sup>2</sup>, Shiping Chen<sup>3</sup>, David Levy<sup>4</sup>, Jicheng You<sup>5</sup>, Danping Yan<sup>\*</sup>

<sup>1</sup> School of Software Engineering, Huazhong University of Science and Technology, Wuhan, 430074, China

<sup>2</sup> College of Public Administration, Huazhong University of Science and Technology, Wuhan, 430074, China

<sup>3</sup> Data61, Commonwealth Scientific and Industrial Research Organization, Marsfield, 2122, Australia

<sup>4</sup> School of Electrical and Information Engineering, University of Sydney, Sydney, 2006, Australia

<sup>5</sup> School of Software Engineering, Huazhong University of Science and Technology, Wuhan, 430074, China

## ABSTRACT

Maximum likelihood estimation is a method of estimating the parameters of a statistical model in statistics. It has been widely used in a good many multi-disciplines such as econometrics, data modelling in nuclear and particle physics, and geographical satellite image classification, and so forth. Over the past decade, although many conventional numerical approximation approaches have been most successfully developed to solve the problems of maximum likelihood parameter estimation, bio-inspired optimization techniques have shown promising performance and gained an incredible recognition as an attractive solution to such problems. This review paper attempts to offer a comprehensive perspective of conventional and bio-inspired optimization techniques in maximum likelihood parameter estimation so as to highlight the challenges and key issues and encourage the researches for further progress.

**Keywords:** *maximum likelihood estimation; bio-inspired optimization; differential evolution; swarm intelligence-based algorithm; genetic algorithm; particle swarm optimization; ant colony optimization.*

## 1 Introduction

Parameter estimation is of very important significance to identifying systematic static and dynamic models since it is known that these indispensable parameters are often deterministic of the systematic probabilistic certainty properties. Maximum likelihood method is on the basis of probabilistic statistics elements. This twofold leads to forming maximum likelihood parameter estimation which is deemed as one of classically probabilistic Bayesian methods. Maximum likelihood estimation has attractive limiting properties of consistency, asymptotic normality and efficiency. It is widely utilized in plenty of domains like dynamic parameter identification in aircrafts, error coefficient estimation in guidance inertia instruments and flow surveillance in traffic engineering. Till date, several optimization methods have been proposed to solve the problems of maximum likelihood parameter estimation, which can be broadly divided into three categories, conventional parsing methods, conventional numerical approximation and bio-inspired optimization approaches.

Nowadays, maximum likelihood parameters in static systems are generally estimated by conventional parsing methods. However, since most of the problems of maximum likelihood parameter estimation in dynamic systems are highly nonlinear, they are difficult to solve by conventional parsing method. As a result, researchers tend to seek for conventional numerical approximation techniques to overcome the difficulties. Among these conventional techniques,

### ARTICLE INFO

Received: April 22, 2018

Accepted: May 2, 2018

Available online: October 23, 2018

### \*CORRESPONDING AUTHOR

Danping Yan, College of Public Administration, Huazhong University of Science and Technology, Wuhan, 430074, China E-mail:

[danpingyan@hust.edu.cn](mailto:danpingyan@hust.edu.cn)

### CITATION

Lu Y, Zhou M, Chen S, Levy D, You J and Yan D. A Perspective of Conventional and Bio-inspired Optimization Techniques in Maximum Likelihood Parameter Estimation. Journal of Autonomous Intelligence 2018; 1(2): 22. doi: 10.32629/jai.v1i2.22

### COPYRIGHT

Copyright © 2018 by author(s) and Frontier Scientific Publishing. This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC4.0).  
<https://creativecommons.org/licenses/by-nc/4.0/>

the typical ones are Gauss-Newton or Newton-Raphson algorithm, Lagrange multipliers algorithm, recursive or iterative least squares methods and Monte Carlo expectation maximization algorithm.

Over the past few decades, modern bio-inspired optimization techniques which are the stochastic search methods inspired by the biological evolutionary nature, are becoming much popular in maximum likelihood parameter estimation. They can be grouped into several categories, namely, swarm intelligence based algorithms, evolution based algorithms and neural network based algorithms. More recently, as they have shown promising results in estimating the maximum likelihood parameters, they have been promptly progressed.

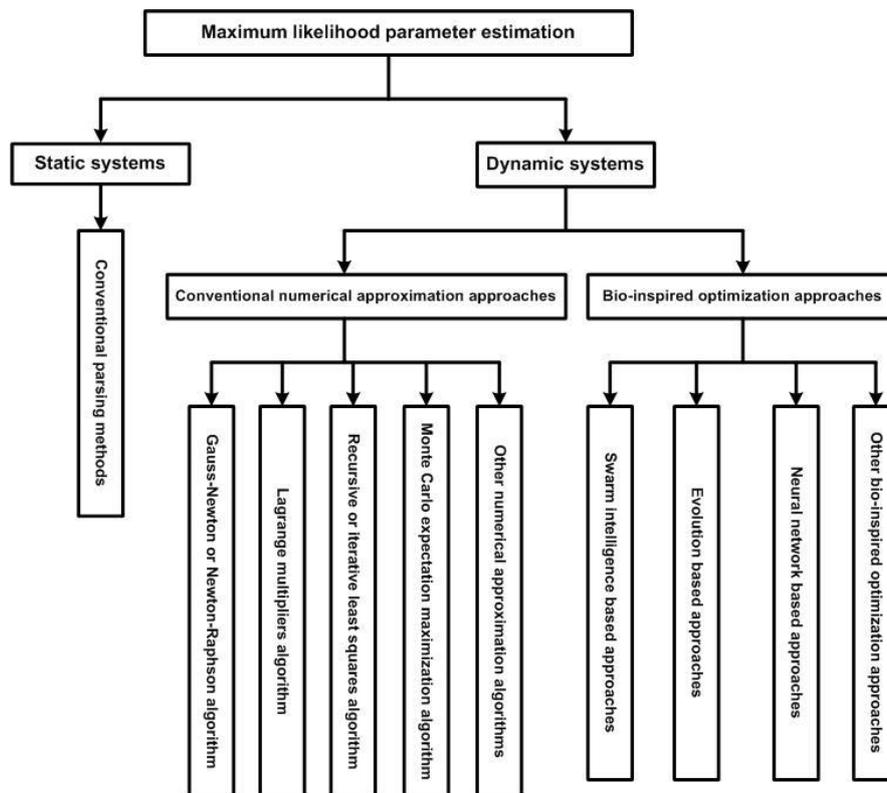
The taxonomy of the conventional parsing, numerical approximation and bio-inspired optimization approaches is presented in Figure 1. Surrounding the taxonomic category, we give a detailed in-depth investigation of these techniques so as to propel the march of maximum likelihood

parameter estimation. The remainder of the paper is arranged below. Section 2 presents the description and formulation of maximum likelihood parameter estimation of static and dynamic systems. Section 3 describes the conventional numerical approximation techniques. Section 4 depicts the bio-inspired optimization techniques. Finally, conclusions and future work are discussed in Section 5.

## 2 Description and formulation of maximum likelihood parameter estimation of static and dynamic systems

### 2.1 Maximum likelihood functions of static systems

In practice, in accordance with stochastic observed sequences in static systems which are either independent or serial, maximum likelihood functions can be classified into two categories, independent observation, serial observation.



**Figure 1:** The taxonomy of the conventional parsing, numerical approximation and bio-inspired optimization approaches.

If  $y(1), y(2), \dots, y(N)$  is a set of stochastic independent observed sequence and  $\theta$  is an

unknown parameter, then there exists a likelihood function

$$L(Y_N | \theta) = p(y(1) | \theta) p(y(2) | \theta) \cdots p(y(N) | \theta) = \prod_{i=1}^N p(y(i) | \theta). \quad (1)$$

If  $y(1), y(2), \dots, y(N)$  is a set of stochastic serial observed sequence and  $\theta$  is an unknown parameter, then there exists a likelihood function

$$L(Y_N | \theta) = p(y(N), y(N-1), \dots, y(2), y(1) | \theta). \quad (2)$$

The maximum likelihood function can be expressed by the following equation

$$\text{Max}_\theta L(Y_N | \theta) = \text{Max}_\theta (L(Y_N | \theta)). \quad (3)$$

## 2.2 Maximum likelihood functions of dynamic systems

The typical model of general linear dynamic systems is shown in Figure 2.

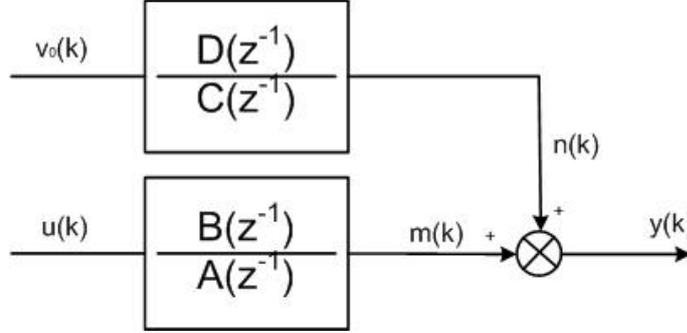


Figure 2: The schematic model of general linear dynamic system.

In Figure 2, for the convenience of computation, assume that there is

$$A(z^{-1}) = C(z^{-1}).$$

As a result, the system can be expressed as

$$A(z^{-1})y(k) = B(z^{-1})u(k) + D(z^{-1})v_0(k), \quad (4)$$

where  $u(k)$  is the system input,  $y(k)$  is the system output, and  $v_0(k)$  is a white noise sequence with the normal distribution, i.e.  $v_0(k) \sim N(0, \sigma^2)$ ,  $A(z^{-1})$ ,  $B(z^{-1})$  and  $D(z^{-1})$  are the polynomials in the shift operator  $z^{-1}$  whilst they are presented below

$$\begin{cases} A(z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_{n_a} z^{-n_a} \\ B(z^{-1}) = b_1 z^{-1} + b_2 z^{-2} + \cdots + b_{n_b} z^{-n_b} \\ A(z^{-1}) = 1 + d_1 z^{-1} + d_2 z^{-2} + \cdots + d_{n_d} z^{-n_d} \end{cases} \quad (5)$$

The inner variables  $m(k)$  and  $n(k)$  are unmeasurable. Assume that the system is asymptotically stable, that the degrees  $n_a$ ,  $n_b$  and  $n_d$  are known, and  $y(k) = 0$ ,  $u(k) = 0$ , and  $v_0(k) = 0$  for  $k \leq 0$ .

Define the parameter vectors  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{d}$ ,  $\boldsymbol{\theta}$ ,  $\mathbf{Y}(k)$ ,  $\mathbf{U}(k)$  and  $\mathbf{V}_0(k)$  as

$$\mathbf{a} := [a_1, a_2, \dots, a_{n_a}]^T \in \mathbb{R}^{n_a},$$

$$\mathbf{b} := [b_1, b_2, \dots, b_{n_b}]^T \in \mathbb{R}^{n_b},$$

$$\mathbf{d} := [d_1, d_2, \dots, d_{n_d}]^T \in \mathbb{R}^{n_d},$$

$$\boldsymbol{\theta} := \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{d} \end{bmatrix} \in \mathbb{R}^{n_a + n_b + n_d},$$

$$\mathbf{Y}(k) = [y(k-1), y(k-2), \dots, y(k-n_a)]^T \in \mathbb{R}^{n_a},$$

$$\mathbf{U}(k) = [u(k-1), u(k-2), \dots, u(k-n_b)]^T \in \mathbb{R}^{n_b},$$

$$\mathbf{V}_0(k) = [v_0(k-1), v_0(k-2), \dots, v_0(k-n_a)]^T \in \mathbb{R}^{n_d},$$

where the superscript T denotes the matrix transpose. Accordingly, the system expression (1) is also considered as

$$y(k) = -\mathbf{a}^T \mathbf{Y}(k) + \mathbf{b}^T \mathbf{U}(k) + \mathbf{d}^T \mathbf{V}_0(k) + v_0(k). \quad (6)$$

For a given set of measurements  $\mathbf{u}_N := [u(1), u(2), \dots, u(N)]$  and  $\mathbf{y}_N := [y(1), y(2), \dots, y(N)]$ , let the likelihood function  $L(\mathbf{y}_N | \mathbf{u}_{N-1}, \theta)$  equal the probability density function  $p(\mathbf{y}_N | \mathbf{u}_{N-1}, \theta)$ .

The likelihood function  $L(\mathbf{y}_N | \mathbf{u}_{N-1}, \theta)$  can be expressed as

$$L(\mathbf{y}_N | \mathbf{u}_{N-1}, \theta) = p(\mathbf{y}_N | \mathbf{u}_{N-1}, \theta) = \prod_{k=1}^N p(-\mathbf{a}^T \mathbf{Y}(k) + \mathbf{b}^T \mathbf{U}(k) + \mathbf{d}^T \mathbf{V}_0(k) + v_0(k) | \mathbf{y}_{N-1}, \mathbf{u}_{N-1}, \theta). \quad (7)$$

The maximum likelihood function can be expressed by the following equation

$$\text{Max}_\theta L(\mathbf{y}_N | \mathbf{u}_{N-1}, \theta) = \text{Max}_\theta (L(\mathbf{y}_N | \mathbf{u}_{N-1}, \theta)). \quad (8)$$

## 3 Conventional numerical approximation techniques

### 3.1 Gauss-Newton or Newton-Raphson algorithm

Either Gauss-Newton or Newton-Raphson algorithm [Li and Ding (2011); Stoica and Babu (2011); Wang et al. (2012); Haryanto and Hong (2013); Hu and Liu (2013); Dosiek et al. (2013); Mohsin et al. (2013); Yao et al. (2014)] is a modification of Newton's method for finding successively a minimum or better approximations to the roots of real-valued functions. They are often used to solve non-linear least squares problems.

Li and Ding (2011) presented a maximum likelihood multi-innovation stochastic gradient algorithm for the identification problems of Hammerstein finite impulse response moving average systems.

Stoica and Babu (2011) introduced a maximum-likelihood method for the nonparametric estimation of smooth spectra from irregularly sampled observations by using the Newton-Raphson method.

Wang et al. (2012) developed a new penalized maximum likelihood cost function by transforming the variables of target position and bearings. They presented a location-penalized maximum likelihood estimator for bearing only target localization on the basis of the Quasi-Newton algorithm.

Haryanto and Hong (2013) applied the ML method for the parameter identification of SISO- and MISO-Wiener-Hammerstein model. And they proposed an iterative gradient-based search approach for the maximum likelihood estimation.

Hu and Liu (2013) proposed a recursive maximum likelihood method based on interior-point algorithm to online estimate the uncertain aerodynamic parameters for hypersonic vehicles by the Gauss-Newton method.

Dosiek et al. (2013) presented a method of directly estimating the variance of each mode estimate in addition to estimating the frequency and damping of each mode in an online setting using a recursive maximum likelihood estimator. The

variance estimates are achieved using two closed-form multidimensional Taylor series approximations, the details of which are fully derived.

In Mohsin et al. (2013), concentration of acidity and major ions in the rainfall in UK is analyzed by assuming a bivariate pseudo-Gamma distribution. The model parameters are estimated by using the maximum likelihood method and the goodness of fit is checked. Furthermore, the non-informative Jeffreys prior for the distribution parameters is derived and a hybrid Gibbs sampling strategy is proposed to sample the corresponding posterior for conducting an objective Bayesian analysis. Finally, related quantities such as the deposition flux density are derived where the general pattern of the observed data appears to follow the fitted densities closely.

Yao et al. (2014) proposed a novel nonlinear feature bias adaptation compensation framework for large vocabulary speech recognition exploiting the core ideas in extreme learning machine. More specifically, they estimated a nonlinear time-dependent bias using an extreme learning machine-like single-hidden-layer neural network to compensate for the mismatch between the training and testing data. In the single-hidden-layer neural network, the lower-layer weights are randomly generated and the output layer contains only linear units just as in the extreme learning machine. Different from extreme learning machine, however, the single-hidden-layer neural network is optimized to maximize the likelihood of either the transformed feature, for which a closed-form solution is derived, or the observed feature, for which an efficient second-order Gauss-Newton method is developed.

### 3.2 Lagrange multipliers algorithm

Lagrange multipliers algorithm [Baltagi and Bresson (2011); Pan et al. (2014); Ogasawara (2016)] is a method for finding the local maxima and minima of a function subject to equality constraints in the domain of mathematical optimization.

Baltagi and Bresson (2011) proposed maximum likelihood estimators for panel seemingly unrelated regressions with both spatial lag and spatial error

components. They studied the general case where spatial effects are incorporated via spatial errors terms and via a spatial lag dependent variable and where the heterogeneity in the panel is incorporated via an error component specification. They proposed joint and conditional Lagrange multiplier tests for spatial autocorrelation and random effects for this spatial SUR panel model. The small sample performance of the proposed estimators and tests are examined using Monte Carlo experiments. An empirical application to hedonic housing prices in Paris illustrate these methods. The proposed specification uses a system of three SUR equations corresponding to three types of flats within 80 districts of Paris over the period 1990-2003. They tested for spatial effects and heterogeneity and find reasonable estimates of the shadow prices for housing characteristics.

Pan et al. (2014) proposed a systematic regularization optimization approach considering a Lagrangian dual relaxation of the maximum likelihood detection problem.

In Ogasawara (2016), inverse expansions of parameter estimators are given in terms of their true values, where the estimators are obtained by the maximum likelihood and weighted score methods with constraints placed on the parameters using Lagrange multipliers. The corresponding expansions for estimated Lagrange multipliers are also given. These expansions are derived before and after studentization. The results with studentization give one-sided confidence intervals for the parameters up to third-order accuracy. As an application of the weighted score method, a modified Jeffreys prior to remove the asymptotic biases of the Lagrange multipliers as well as the parameter estimators is obtained under canonical parametrization in the exponential family.

### 3.3 Recursive or iterative least squares algorithm

Recursive or iterative least squares algorithm [Wang et al. (2010); Li et al. (2012); Stathakis et al. (2014); Dutta et al. (2014)] is considered as an online approach to solving the least squares problems by a

given iteration algorithm.

Wang et al. (2010) dealt with the parameter identification problem of Hammerstein output error auto-regressive systems with different nonlinearities by combining the key-term separation principle and the auxiliary model identification idea. They used the key-term separation principle to present auxiliary model based recursive generalized least squares algorithms.

Li et al. (2012) presented a recursive maximum likelihood least squares identification method for input nonlinear finite impulse response moving average systems.

Stathakis et al. (2014) presented an affinely modified version of the ML estimation that uniformly outperforms over all SNR values of the traditional maximum likelihood estimation in terms of the mean-square errors in closed-form.

Dutta et al. (2014) applied TRANUS to calibrate the land-use module of the Grenoble model, based on the mean-square errors.

### 3.4 Monte Carlo expectation maximization algorithm

Monte Carlo expectation maximization algorithm [Denoeux (2011); Liu et al. (2012); Augustyniak (2014); Martin and Hanb (2016); Galimberti and Soffritti (2014); Wang et al. (2015); González et al. (2015); Sun et al. (2016)] is an iterative method for finding maximum likelihood or maximum a posteriori estimates of parameters in statistical models, where the model depends on unobserved latent variables. The expectation maximization iteration alternates between a log-likelihood parameter estimation expectation step and a log-likelihood parameter estimation maximization step. These estimated parameters are then used to determine the distribution of the latent variables in the next log-likelihood parameter estimation expectation step.

Denoeux (2011) used the expectation maximization algorithm to estimate the parameters in a parametric statistical model when the observations are fuzzy and are assumed to be related to underlying crisp realizations of a random sample.

Likewise, Liu et al. (2012) proposed an efficient maximum likelihood direction-of-arrival estimator based on a spatially over-complete array output formulation, where the Monte Carlo expectation maximization algorithm is exploited.

Based on both the Monte Carlo expectation maximization algorithm and the sampling importance, Augustyniak (2014) developed a novel approach to calculate the maximum likelihood estimator and asymptotic variance covariance matrix of the Markov switching GARCH model.

To avoid specification of a particular distribution for the error in a regression model, Martin and Hanb (2016) proposed a flexible scale mixture model with a nonparametric mixing distribution. They exploited hybrid predictive recursion expectation maximization algorithm to solve this model.

Galimberti and Soffritti (2014) introduced a novel approach that assumes that the error terms follow a finite mixture of  $t$  distributions. This assumption allows for an extension of multivariate linear regression models, making these models more versatile and robust against the presence of outliers in the error term distribution. The issues of model identifiability and maximum likelihood estimation are addressed. Identifiability conditions are provided and an expectation maximization algorithm for estimating the model parameters is developed.

The pyramidal dual-tree directional filter bank transform is a new image decomposition. Wang et al. presented a new color image segmentation algorithm based on pyramidal dual-tree directional filter bank domain hidden Markov tree model. A color image segmentation using pyramidal dual-tree directional filter bank domain hidden Markov tree model is developed, in which expectation maximization parameter estimation, Wang et al. (2015) multi-scale raw segmentation, context based multi-scale fusion, and majority-vote based color component fusion are used.

In González et al. (2015), the maximum likelihood estimation of the parameters of interest for the controlled branching process is addressed under various sample schemes. Firstly, assuming that the entire family tree can be observed, the corresponding

estimators are obtained and their asymptotic properties investigated. Secondly, since in practice it is not usual to observe such a sample, the maximum likelihood estimation is initially considered using the sample given by the total number of individuals and progenitors of each generation, and then using the sample given by only the generation sizes. Expectation maximization algorithms are developed to address these problems as incomplete data estimation problems.

Sun et al. (2016) presented a recursive maximum likelihood state estimator based on the expectation maximization algorithm for Markov jump linear systems with uncertain mode-dependent delays. To calculate the posterior probability of each possible candidate time delay, a recursive algorithm is derived within the Bayesian framework conditioned on the likelihood density function of state with respect to the candidate time delay related to the reference mode. By combining the optimal principle of expectation maximization and the interacting multiple mode approximation, they proposed a double-reweight interacting multiple model algorithm to obtain the maximum a posterior estimator of state that is of low computational complexity.

### 3.5 Other numerical approximation algorithms

Besides the aforementioned methods, there are some other numerical approximation algorithms to be worth noting.

Wen and Chen (2011) developed a very efficient and stable computation algorithm for nonparametric maximum likelihood estimation of gamma-frailty Cox models with clustered current status data, based on a set of self-consistency equations and the contraction principle.

Baghishani and Mohammadzadeh (2011) proposed a computationally efficient strategy to fit spatial generalized linear mixed models, based on the data cloning method suggested by Lele et al. in 2007 instead of the Bayesian approach.

Pence et al. (2011) combined polynomial chaos theory with maximum likelihood estimation for a

novel approach to recursive parameter estimation in state space systems.

Hirai and Yamanishi (2013) addressed the issue of estimating the number of mixture components for a Gaussian mixture model from a given data sequence. They computed the normalized maximum likelihood code length for the data sequence relative to a Gaussian mixture model, then to find the mixture size that attains the minimum of the normalized maximum likelihood on the basis of the minimum description length principle.

In settings where high-level inferences are made based on registered image data, the registration uncertainty can contain important information. Risholm et al. (2013) proposed a Bayesian non-rigid registration framework where conventional dissimilarity and regularization energies can be included in the likelihood and the prior distribution on deformations respectively through the use of Boltzmann's distribution. The posterior distribution is characterized using Markov chain Monte Carlo methods with the effect of the Boltzmann temperature hyper-parameters marginalized under broad uninformative hyper-prior distribution. The Monte Carlo chain permits estimation of the most likely deformation as well as the associated uncertainty. On synthetic examples, they demonstrated the ability of the method to identify the maximum a posteriori estimate and the associated posterior uncertainty, and demonstrated that the posterior distribution can be non-Gaussian.

A new modal parameter estimation method to directly establish modal models of structural dynamic systems satisfying two physically motivated constraints is presented. The constraints imposed in the identified modal model are the reciprocity of the frequency response functions and the estimation of normal real modes. The motivation behind the first constraint comes from the fact that modal analysis theory shows that the frequency response function matrix and therefore the residue matrices are symmetric for non-gyroscopic, non-circulatory, and passive mechanical systems. Therefore, El-Kafafy et al. (2016) proposed a constrained maximum likelihood modal parameter estimation method and successfully applied it to two real experimental

data-sets measured on fully trimmed cars.

Skaug and Yu (2014) used the Laplace approximation to perform maximum likelihood estimation of univariate and multivariate stochastic volatility models. The implementation of the Laplace approximation is greatly simplified by the use of a numerical technique known as automatic differentiation.

## **4 Bio-inspired optimization techniques**

### **4.1 Swarm intelligence based approaches**

Swarm intelligence based approaches are based on the collective behaviors of decentralized, self-organized natural systems such as ant colonies, bird flocking, animal herding, bacterial growth, fish schooling, microbial intelligence, and so forth. The emergence of intelligent behaviors leads to various swarm intelligence algorithms for difficult types of optimization problems. The typically proposed algorithms are particle swarm optimization, antcolony optimization, artificial bee algorithm, artificial bee colony algorithm, bat algorithm, bacterial foraging, artificial immune systems and firefly algorithm. They are used to seek for maximum likelihood or maximum a posteriori estimates of parameters in statistical models.

An et al. (2012) presented a novel parametrization for arbitrary covariance matrices that allow independent updating of individual parameters while retaining validity of the resultant matrices. They used particle swarm optimization based parameter matching technique to mitigate the issues related with the existence of multiple candidate solutions that are equivalent under the permutation of the components in Gaussian mixture models.

Chang and Chen (2013) proposed a modified particle swarm optimization approach to compute the maximum likelihood functions and find the global minimum cost function for array calibration.

Tay et al. (2013) proposed a novel algorithm called the evolutionary data-conscious AIRS algorithm that accentuates and capitalizes on 3 additional immune mechanisms observed from the natural immune system. These mechanisms are associated to the phenomena exhibited by the

antibodies in response to the concentration, location and type of foreign antigens. Bio-mimicking these observations empower evolutionary data-conscious AIRS algorithm with the ability to robustly adapt to the different density, distribution and characteristics exhibited by each data class. This provides competitive advantages for the algorithm to better characterize and learn the underlying pattern of the data.

To reduce the high computational burden of maximum likelihood method and make it more suitable to engineering applications, Zhang et al. (2013) applied the artificial bee colony algorithm to maximize the likelihood function for direction-of-arrival estimation. The algorithm is originally used to optimize multi-variable functions by imitating the behavior of bee colony finding excellent nectar sources in the nature environment. It offers an excellent alternative to the conventional methods in maximum likelihood method-direction-of-arrival estimation. The performance of artificial bee colony-based maximum likelihood and other popular meta-heuristic-based maximum likelihood methods for direction-of-arrival estimation are compared for various scenarios of convergence, signal-to-noise ratio, and number of iterations. The computation loads of artificial bee colony-based maximum likelihood and the conventional maximum likelihood methods for direction-of-arrival estimation are also investigated.

Giacomina et al. (2015) proposed a probabilistic approach for designing nonlinear optimal robust tracking controllers for unmanned aerial vehicles. The controller design is formulated in terms of a multi-objective optimization problem that is solved by using a modified particle swarm optimization algorithm, offering high likelihood of finding an optimal or near-optimal global solution.

Inspired by ant's stochastic behavior in search for multiple food sources, Xu et al. (2015) proposed a cooperating multi-task ant system for tracking multiple synthetic objects as well as multiple real cells in a bio-medical field. In their framework, each ant colony is assumed and assigned to fulfill a given task to estimate the state of an object. Furthermore, two ant levels are used, i.e., ant individual level and

ant cooperation level. In the ant individual level, ants within one colony perform independently, and the motion of each individual is probabilistically determined by both its intended motion modes and the likelihood function score. In the ant cooperation level, each ant adjusts individual state within its influence region according to heuristic information of all other ants within the same colony, while the global best template at current iteration is found among all ant colonies and utilized to update ant model probability, influence region, and probability of fulfilling task.

## 4.2 Evolution based approaches

Evolution based approaches follow the evolutionary principles of Charles Darwin's theory of survival of the fittest. They usually use multi-agents or search vectors with mutation and crossover factors to carry out the search. Among these approaches, frequently used ones are genetic algorithm or differential evolution based approaches.

Evolutionary relationships among species can be represented by a phylogenetic tree and inferred by optimizing some measure of fitness, such as the statistical likelihood of the tree (given a model of the evolutionary process and a data set). The combinatorial complexity of inferring the topology of the best tree makes phylogenetic inference ideal for genetic algorithm. Two existing algorithms for phylogenetic inference (neighbor-joining and maximum likelihood) are co-utilized within a genetic algorithm and enable the phenotype and genotype to be assigned quite different representations [Poladian (2005)].

Aguirregarria (2005) presented a hybrid genetic algorithm to obtain maximum likelihood estimates of parameters in structural econometric models with multiple equilibria. The algorithm combines a pseudo maximum likelihood procedure with a genetic algorithm. The genetic algorithm searches globally over the large space of possible combinations of multiple equilibria in the data. The pseudo maximum likelihood procedure avoids the computation of all the equilibria associated with every trial value of the structural parameters.

Obtaining good estimates of structural parameters from observed data is a particularly challenging task owing to the complex likelihood functions that often accompany such problems. As a result, sophisticated optimization routines are typically required to produce maximum likelihood estimates of the desired parameters. Evolutionary algorithms comprise one such approach, whereby nature-inspired mutation and crossover operations allow the sensible exploration of even multi-modal functions, in search of a global maximum. The challenge, of course, is to balance broad coverage in parameter space with the speed required to obtain such estimates. Reed et al. (2013) proposed a modified version of the differential evolution algorithm to address the problem. Their idea is to adjust both mutation and cross-over rates, during the optimization, in a manner that increases the convergence rate to the desired solution.

### 4.3 Neural network based approaches

Neural network based approaches mainly depend on a family of models inspired by biological neural networks to estimate or approximate the maximum likelihood parameter estimation functions. Their distinct properties are both interconnected neurons to exchange messages between each other and their adaptive connections with numeric weights.

In Jelonek and Reilly (1990), a neural network-based system for a maximum-likelihood estimation of directions of arrival is described. A novel analog neural network implementation of the maximum-likelihood algorithm is presented. Properties of the neural network are discussed with respect to stability and convergence. The performance and behavioral simulations of the network's dynamics are presented.

Firmino et al. (2014) introduced a two-step method for correcting and combining forecasting models. Firstly, the stochastic process underlying the bias of each predictive model is built according to a recursive ARIMA algorithm in order to achieve a white noise behavior. At each iteration of the algorithm the best ARIMA adjustment is determined according to a given information criterion. Then, in

the light of the corrected predictions, it is considered a maximum likelihood combined estimator. Applications involving single ARIMA and artificial neural networks models for Dow Jones Industrial Average Index, S\&P500 Index, Google Stock Value, and Nasdaq Index series illustrate the usefulness of the proposed framework.

Clinical risk likelihood prediction is important for individuals at risk of cardiovascular disease given the fact that it is the leading causes of death in many developed countries. To this end, Tay et al. (2015) introduced a novel learning algorithm to develop a cardiovascular disease risk prediction tool. This novel neural-inspired algorithm, called the artificial neural cell system for classification, is inspired by mechanisms that develop the brain, empower it with capabilities such as information processing/storage and recall, decision making, and initiate actions on external environment.

### 4.4 Other bio-inspired optimization approaches

In particular, hybrid conventional and bio-inspired optimization approaches are worth noting. They fully take advantages of two types of methods to perform maximum likelihood estimation.

Mirikitani and Nikolaev (2011) built on electricity spot prices by proposing an expectation maximization algorithm for maximum likelihood estimation of recurrent neural networks utilizing the Kalman filter and smoother. This involves inference of both parameters and hyper-parameters of the model which takes into account the model uncertainty and noise in the data. The expectation maximization algorithm uses a forward filtering and backward smoothing expectation step, followed by a hyper-parameter estimation maximization step. The model is validated across two data sets of different power exchanges. It is found that after learning a posterior hyper-parameters, the proposed algorithm outperforms the real-time recurrent learning and the extended Kalman Filtering algorithm for recurrent networks, as well as other contemporary models that have been previously applied to the modeling of electricity spot prices.

## 5 Outlook and future work

The focus of the paper is the investigation of the recent use of conventional and bio-inspired optimization techniques in maximum likelihood estimation. Among the conventional numerical approximation techniques, a large percent of research works are related to Gauss-Newton or Newton-Raphson and Monte Carlo expectation maximization algorithms. Meanwhile, swarm intelligence based and evolution based approaches are more popular in the bio-inspired optimization techniques. Especially, ant colony optimization, particle swarm optimization and differential evolution have gained more and more attention from the researchers because they all have comparatively better global search performance, compared with the conventional numerical approximation techniques.

Despite the success of the recent use of conventional and bio-inspired optimization techniques in maximum likelihood estimation, many challenging problems still remain unsolved. In this context, it is typical that the considered models are nonlinear due to the nonlinearity of the behavior of the modeled systems. On the other hand, these models have many parameters or are of high dimensionality. Moreover, the measurements are imperfect due to measurement noise and partial observation. Additionally, the conventional numerical approximation techniques are mostly attributed to the local derivative-based methods. They require that the optimization problems should be differentiable and asymptotic. The Monte Carlo expectation maximization algorithm easily gets trapped in a local maximum as the objective being a non-concave optimization problem, and is greatly affected by the initialization conditions. However, the bio-inspired optimization techniques do not use the gradients of the problems being optimized. These properties make maximum likelihood parameter estimation a challenging optimization problem, calling for the development of advanced optimization methods.

Further work is needed to develop the advanced bio-inspired optimization techniques adaptive to maximum likelihood parameter estimation of

complex modeled systems with high nonlinearity and dimensionality, i. e., bio-inspired optimization approaches for large scale nonlinear problems. In the meantime, of course, we need to further improve the conventional numerical approximation techniques in maximum likelihood parameter estimation. Last, but not least, we yet need to develop the hybrid conventional and bio-inspired techniques to be suitable for estimating maximum likelihood parameters in light of their mutual merits.

## Acknowledgments

This work is partly supported by the Fundamental Research Funds for the Central Universities in China (HUST: 2016YXMS105). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

## References

1. Aguirregariria, V., A hybrid genetic algorithm for the maximum likelihood estimation of models with multiple equilibria: A first report, *New Mathematics and Natural Computation* 1 (2) (2005) 295-303.
2. An, ç, ., Aksoy, S., Ankan, O., Maximum likelihood estimation of Gaussian mixture models using stochastic search, *Pattern Recognition* 45 (7) (2012) 2804-2816.
3. Augustyniak, M., Maximum likelihood estimation of the Markov-switching GARCH model, *Computational Statistics and Data Analysis* 76 (2) (2014) 61-75.
4. Baghishani, H., Mohammadzadeh, M., A data cloning algorithm for computing maximum likelihood estimates in spatial generalized linear mixed models, *Computational Statistics and Data Analysis* 55 (4) (2011) 1748-1759.
5. Baltagi, B. H., Bresson, G., Maximum likelihood estimation and Lagrange multiplier tests for panel seemingly unrelated regressions with spatial lag and spatial errors: An application to hedonic housing prices in Paris, *Journal of Urban Economics* 69 (1) (2011)

- 24-42.
6. Chang, J.-C., Chen, C.-Y., Maximum likelihood DOA estimation with sensor position perturbation using particle swarm optimization, *Advances in Intelligent Systems and Applications: Smart Innovation, Systems and Technologies* 20 (1), Springer-Verlag, Berlin, Heidelberg, 2013, pp. 187-197.
  7. Denoeux, T., Maximum likelihood estimation from fuzzy data using the EM algorithm, *Fuzzy Sets and Systems* 183 (1) (2011) 72-91.
  8. Dosiek, L., Pierre, J. W., Follum, J., A recursive maximum likelihood estimator for the online estimation of electromechanical modes with error bounds, *IEEE Transactions on Power Systems* 28 (1) (2013) 441-451.
  9. Dutta, P., Arnaud, E., Prados, E., Saujot, M., Calibration of an integrated land-use and transportation model using maximum likelihood estimation, *IEEE Transactions on Computers* 63 (1) (2014) 167-178.
  10. El-Kafafy, M., Peeters, B., Guillaume, P., Troyer, T. D., Constrained maximum likelihood modal parameter identification applied to structural dynamics, *Mechanical Systems and Signal Processing* 72-73 (2016) 567-589.
  11. Firmino, P. R. A., Mattos Netob, P. S. G. d., Ferreira, T. A. E., Correcting and combining time series forecasters, *Neural Networks* 50 (2014) 1-11.
  12. Haryanto, A., Hong, K.-S., Maximum likelihood identification of Wiener-Hammerstein models, *Mechanical Systems and Signal Processing* 41 (1-2) (2013) 54-70.
  13. Hirai, S., Yamanishi, K., Efficient computation of normalized maximum likelihood codes for Gaussian mixture models with its applications to clustering, *IEEE Transactions on Information Theory* 59 (11) (2013) 7718-7727.
  14. Hu, C., Liu, Q., Online identification for hypersonic vehicle using recursive maximum likelihood method based on interior-point algorithm, in: *25th Chinese Control and Decision Conference*, Guiyang, China, IEEE, 2013, pp. 1862-1867.
  15. Galimberti, G., Soffritti, G., A multivariate linear regression analysis using finite mixtures of t distributions, *Computational Statistics and Data Analysis* 71 (2014) 138-150.
  16. Giacomina, P. A. S., Hemerlyb, E. M., Pedrycz, W., A probabilistic approach for designing nonlinear optimal robust tracking controllers for unmanned aerial vehicles, *Applied Soft Computing* 34 (2015) 26-38.
  17. González, M., Minuesa, C., Puerto, I. d., Maximum likelihood estimation and expectation-maximization algorithm for controlled branching processes, *Computational Statistics and Data Analysis* 93 (2016) 209-227.
  18. Jelonek, T. M., Reilly, J. P., Maximum likelihood estimation for direction of arrival using a nonlinear optimising neural network, in: *International Joint Conference on Neural Networks*, IEEE, San Diego, 1990, pp. 253-258.
  19. Li, J., Ding, F., Maximum likelihood stochastic gradient estimation for Hammerstein systems with colored noise based on the key term separation technique, *Computers and Mathematics with Applications* 62 (11) (2011) 4170-4177.
  20. Li, J., Ding, F., Yang, G., Maximum likelihood least squares identification method for input nonlinear finite impulse response moving average systems, *Mathematical and Computer Modelling* 55 (3-4) (2012) 442-450.
  21. Liu, Z.-M., Huang, Z.-T., Zhou, Y.-Y., An efficient maximum likelihood method for direction-of-arrival estimation via sparse Bayesian learning, *IEEE Transactions on Wireless Communications* 11 (10) (2012) 1-11.
  22. Martin, R., Hanb, Z., A semiparametric scale-mixture regression model and predictive recursion maximum likelihood, *Computational Statistics and Data Analysis* 94 (2016) 75-85.
  23. Mirikitani, D., Nikolaev, N., Nonlinear maximum likelihood estimation of electricity spot prices using recurrent neural networks, *Neural Computing and Applications* 20 (1) (2011) 79-89.
  24. Mohsin, M., Kazianka, H., Pilz, J., Likelihood and objective Bayesian modeling of acidity and major ions in rain fall using a bivariate

- pseudo-Gamma distribution, *Computers & Geosciences* 54 (2013) 269-278.
25. Ogasawara, H., Asymptotic expansions for the estimators of Lagrange multipliers and associated parameters by the maximum likelihood and weighted score methods, *Journal of Multivariate Analysis* 147 (2016) 20-37.
  26. Pan, J., Ma, W.-K., Jalden, J., MIMO detection by Lagrangian dual maximum-likelihood relaxation: reinterpreting regularized lattice decoding, *IEEE Transactions on Signal Processing* 62 (2) (2014) 511-524.
  27. Pence, B.L., Fathy, H.K., Stein, J.L., Recursive maximum likelihood parameter estimation for state space systems using polynomial chaos theory, *Automatica* 47 (11) (2011) 2420-2424.
  28. Poladian, L., A GA for maximum likelihood phylogenetic inference using neighbour-joining as a genotype to phenotype mapping, in: *Proceedings of the 7th annual conference on Genetic and evolutionary computation*, New York, USA, ACM, 2005, pp. 415-422.
  29. Reed, H. M., Nichols, J. M., Earls, C. J., A modified differential evolution algorithm for damage identification in submerged shell structures, *Mechanical Systems and Signal Processing* 39(1-2) (2013) 396-408.
  30. Risholm, P., Janoos, F., Norton, I., Golby, A. J., Wells III, W. M., Bayesian characterization of uncertainty in intra-subject non-rigid registration, *Medical Image Analysis* 17 (2013) 538-555.
  31. Skaug, H. J., Yu, J., A flexible and automated likelihood based framework for inference in stochastic volatility models, 76 (2014) 642-654.
  32. Stathakis, E., Jalden, J., Rasmussen, L. K., Skoglund, M., Uniformly improving maximum-likelihood SNR estimation of known signals in Gaussian channels, *IEEE Transactions on Signal Processing* 62 (1) (2014) 156-167.
  33. Stoica, P., Babu, P., Maximum-likelihood nonparametric estimation of smooth spectra from irregularly sampled data, *IEEE Transactions on Signal Processing* 59 (12) (2011) 5746-5758.
  34. Sun, Q., Lim, C.-C., Liu, L., Maximum likelihood state estimation for Markov jump systems with uncertain mode-dependent delays, *Journal of the Franklin Institute* 353 (2016) 594-614.
  35. Tay, D., Poh, C. L., Kitney, R. I., An evolutionary data-conscious artificial immune recognition system, in: *Genetic and evolutionary computation conference*, Amsterdam, Netherland, ACM, 2013, pp. 1101-1108.
  36. Tay, D., Poh, C. L., Kitney, R. I., A novel neural-inspired learning algorithm with application to clinical risk prediction, *Journal of Biomedical Informatics*, 54 (2015) 305-314.
  37. Wang, D. Q., Chu, Y. Y., Yang, G. W., Ding, F., Auxiliary model-based recursive generalized least squares parameter estimation for Hammerstein OEAR systems, *Mathematical and Computer Modelling* 52 (1-2) (2010) 309-317.
  38. Wang, X., Sun, W., Wu Z., Yang H., Wang Q., Color image segmentation using PDTDFB domain hidden Markov tree model, *Applied Soft Computing* 29 (2015) 138-152.
  39. Wang, Z., Luo, J.-A., Zhang, X.-P., A novel location-penalized maximum likelihood estimator for bearing-only target localization, *IEEE Transactions on Signal Processing* 60 (12) (2012) 6166-6181.
  40. Wen, C.-C., Chen, Y.-H., Nonparametric maximum likelihood analysis of clustered current status data with the gamma-frailty Cox model, *Computational Statistics and Data Analysis* 55 (2) (2011) 1053-1060.
  41. Xu, B., Lu, M., Ren, Y., Zhua, P., Shi, J., Cheng, D., Multi-task ant system for multi-object parameter estimation and its application in cell tracking, *Applied Soft Computing* 35 (2015) 449-469.
  42. Yao, K., Yu, D., Deng, L., Gong, Y., A fast maximum likelihood nonlinear feature transformation method for GMM-HMM speaker adaptation, *Neurocomputing* 128 (2014) 145-152.
  43. Zhang, Z., Lin, J., Shi, Y., Application of artificial bee colony algorithm to maximum likelihood DOA estimation, *Journal of Bionic Engineering* 10 (2013) 100-109