Systems & Control Letters 115 (2018) 15–21

Contents lists available at ScienceDirect

Systems & Control Letters

journal homepage: www.elsevier.com/locate/sysconle

Multi-step-length gradient iterative algorithm for equation-error type models

Jing Chen [a,*], Feng Ding [a], Yanjun Liu [a], Quanmin Zhu [b]

[a] Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122, PR China
[b] Department of Engineering Design and Mathematics, University of the West of England, Bristol BS16 1QY, UK

ARTICLE INFO

Article history:
Received 10 November 2017
Received in revised form 31 January 2018
Accepted 2 March 2018

Keywords:
Parameter estimation
Gradient algorithm
Multi-step-length
Convergence rate
EET model

ABSTRACT

This letter develops a multi-step-length gradient iterative algorithm for equation-error type (EET) models. The algorithm analysis is based upon the gradient search principle. By applying the Gram–Schmidt procedure, the EET model can be turned into an orthogonal model, in which each parameter in the parameter vector is independent of the other parameters. Then a multi-step-length gradient iterative algorithm is proposed for the orthogonal model, and can estimate the parameters in one iteration. Finally, based on the estimated parameters, the parameter estimates of the EET model can be computed. Different from the traditional gradient iterative algorithm which has slower convergence rates and is sensitive to the initial values of the unknown variables, the method in this letter has quicker convergence rates and is robust to the initial values of the unknown variables. The simulation studies demonstrate the feasibility and effectiveness of the proposed algorithm.

1. Introduction

System identification is essential for controller design if the parameters of the systems are unknown. A robust controller design assumes that the parameters of the system is known in prior [1,2]. Thus, system identification is a firm foundation of controller design. Recently, many parameter estimation methods have been reported in literatures, such as the least squares algorithms [3,4], the gradient iterative (GI) algorithms [5,6] and the maximum likelihood algorithms [7,8]. The GI algorithm which is called line search method is widely used in system identification, that is because the GI algorithm has less computational efforts. The idea of the GI algorithm is to compute a search direction and then to decide how far to move along that direction at each iteration.

Since the neighboring directions of the GI algorithm are orthogonal, the iterations zigzag toward the solution, which leads to the convergence rates of the GI algorithm be slow [9,10]. Generally, there are two ways to increase the convergence rates, one is to change the direction in order to get a better direction. For example, Abbasbandy et al. provided a conjugate gradient method for fuzzy symmetric positive definite system of linear equations [11], in which the conjugate gradient method can obtain an optimal direction at each iteration. Hussu developed a conjugate-gradient method for optimal control problems with undetermined final time, and this method is suitable for the computation of on–off control policies as well [12]. The other way to increase the convergence rates is to improve the length of the step-length. For example, in [13], a modified stochastic gradient algorithm was proposed to increase the convergence rates by introducing a convergence index. In [14], a data filtering based forgetting factor stochastic gradient algorithm was investigated for Hammerstein systems with saturation and preload nonlinearities, the forgetting factor in the step-length can increase the convergence rates of the stochastic gradient algorithm. Although the modified and forgetting factor stochastic gradient algorithms can increase the convergence rates, they also bring vibration when the estimated parameters are close to the true values [15].

All the above modified gradient based algorithms assume that the step-length is a positive constant no matter what the dimension of the unknown parameters is. However, this assumption may be unreasonable when some elements of the initial parameters are far away from the true elements while some other elements are near to the true elements. In this letter, we assume that each element has a direction and a step-length at each iteration, then a multi-step-length gradient iterative (Mul-GI) algorithm is developed for an EET model in which each parameter is independent of the other parameters. The Mul-GI algorithm can estimate the parameters in one iteration and is robust to the initial parameters. Unfortunately, when each parameter is dependent on the other parameters, the Mul-GI algorithm is less effective than the traditional GI algorithm. In this situation, the Gram–Schmidt procedure is introduced to transform the EET model into an orthogonal model

* Corresponding author.
E-mail addresses: chenjing1981929@126.com (J. Chen), fding@jiangnan.edu.cn (F. Ding), yanjunliu_1983@126.com (Y. Liu), quan.zhu@uwe.ac.uk (Q. Zhu).

https://doi.org/10.1016/j.sysconle.2018.03.003
0167-6911/

whose parameters are uncorrelated. Then the parameters of the orthogonal model can be estimated by the Mul-GI algorithm in one iteration. Finally, the unknown parameters of the EET model can be computed based on the estimated parameters of the orthogonal model. The contributions of this letter are summarized as follows.

(1) The Mul-GI algorithm can increase the convergence rates of the GI algorithm (in one iteration).
(2) The Mul-GI algorithm is robust to the initial parameters.
(3) The Mul-GI algorithm can deal with systems with colored noise.

This letter is organized as follows. Section 2 introduces the gradient search principle. Section 3 develops the Mul-GI algorithm for quadratic functions. Section 4 proposes the Mul-GI algorithm for EET models. Section 5 provides an illustrative example. Finally, concluding remarks are given in Section 6.

2. The gradient search principle

Consider the following quadratic function,

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}, \tag{1}$$

where $\boldsymbol{Q}$ is a symmetric and positive definite matrix. In order to get the minimizer $\boldsymbol{x}^*$, at iteration k, assume that the present point is $\boldsymbol{x}^k$. Based on the gradient search principle, one should determine both the direction $\boldsymbol{p}_k$ and the step-length λ_k to get a new point $\boldsymbol{x}^{k+1}$ which is better than $\boldsymbol{x}^k$,

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \lambda_k \boldsymbol{p}_k, \tag{2}$$

$$\boldsymbol{p}_k = -[\boldsymbol{Q}\boldsymbol{x}_k - \boldsymbol{b}], \tag{3}$$

$$\lambda_k = \frac{(\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{Q} - \boldsymbol{b}^{\mathrm{T}})(\boldsymbol{Q}\boldsymbol{x}_k - \boldsymbol{b})}{(\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{Q} - \boldsymbol{b}^{\mathrm{T}})\boldsymbol{Q}(\boldsymbol{Q}\boldsymbol{x}_k - \boldsymbol{b})}, \tag{4}$$

and then $\boldsymbol{x}^{k+1}$ satisfies $f(\boldsymbol{x}^{k+1}) \leqslant f(\boldsymbol{x}^k)$. At each iteration the direction $\boldsymbol{p}_k$ is orthogonal to the next direction $\boldsymbol{p}_{k+1}$, which indicates that the iterations zigzag toward the solution. Thus the gradient algorithm has slower convergence rates, especially when the point $\boldsymbol{x}_k$ is close to the true solution $\boldsymbol{x}^*$. The following theorem gives the convergence rates of the gradient algorithm for the quadratic function (1) [16].

Theorem 1. *When the gradient iterative method with exact line searches (2) is applied to the strongly convex quadratic function (1), then*

$$\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|_{\boldsymbol{Q}}^2 \leqslant \left(\frac{r_n - r_1}{r_n + r_1}\right) \|\boldsymbol{x}^k - \boldsymbol{x}^*\|_{\boldsymbol{Q}}^2,$$

where $\|\boldsymbol{x}\|_{\boldsymbol{Q}}^2 = \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}$, and $0 < r_1 \leqslant r_2 \leqslant \cdots \leqslant r_n$ are the eigenvalues of $\boldsymbol{Q}$.

According to this theorem, one can conclude that convergence is achieved in one iteration when $\boldsymbol{Q}$ is a multiple of the identity matrix. On the other hand, as the condition number r_n/r_1 increases, the convergence degrades. For example, when $\boldsymbol{Q} = \mathrm{diag}[1, 10]$, $\boldsymbol{b} = [0, 0]^{\mathrm{T}}$, and the initial point is $[2, 2]^{\mathrm{T}}$. Based on the GI algorithm, after 8 iterations, one can get the optimal variable $\boldsymbol{x} = [0, 0]^{\mathrm{T}}$.

When the matrix $\boldsymbol{Q}$ is a diagonal positive definite matrix, a question arises: can we develop a novel algorithm which can get the optimal variable quicker than the GI algorithm? This is the focus of this letter.

Table 1
The estimates with different initial points.

Algorithms	Initial points	Iterations	Initial points	Iterations
GI	[2, 2]	8	[10, 10]	10
Mul-GI	[2, 2]	1	[10, 10]	1

3. The multi-step-length gradient iterative algorithm for quadratic functions

In the traditional gradient search principle, only one step-length can be obtained at each iteration no matter what the dimension of $\boldsymbol{x}$ is. The idea of the multi-step-length gradient search principle is to assign a step-length for each element of $\boldsymbol{x}$ at each iteration, in other words, assign a step-length matrix for $\boldsymbol{x}$ at each iteration, e.g., if $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$, one should assign a step-length λ^i, $i = 1, \ldots, n$ for each element x_i.

3.1. Q is a diagonal and positive definite matrix

Rewrite the quadratic function

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}, \tag{5}$$

where the diagonal positive definite matrix $\boldsymbol{Q}$, the vector $\boldsymbol{b}$ and the unknown variable vector $\boldsymbol{x}$ are defined as

$$\boldsymbol{Q} = \mathrm{diag}[l_1, l_2, \ldots, l_n],\ l_i > 0,\ i = 1, \ldots, n,$$
$$\boldsymbol{b} = [b_1, b_2, \ldots, b_n]^{\mathrm{T}},$$
$$\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}.$$

Based on the gradient search principle, the direction $\boldsymbol{p}_k$ at iteration k can be computed as

$$\boldsymbol{p}_k = -[l_1x_1^k - b_1, l_2x_2^k - b_2, \ldots, l_nx_n^k - b_n]^{\mathrm{T}}. \tag{6}$$

Then $\boldsymbol{x}^{k+1}$ can be expressed as

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - \boldsymbol{\alpha}^k[l_1x_1^k - b_1, l_2x_2^k - b_2, \ldots, l_nx_n^k - b_n]^{\mathrm{T}}, \tag{7}$$

$$\boldsymbol{\alpha}^k = \mathrm{diag}[\lambda_1, \lambda_2, \ldots, \lambda_n]. \tag{8}$$

$\boldsymbol{\alpha}^k$ is a step-length matrix.

In order to obtain λ_i, $i = 1, \ldots, n$, taking the first-order functional derivative of $f(\boldsymbol{x}^k - \boldsymbol{\alpha}^k[l_1x_1^k - b_1, l_2x_2^k - b_2, \ldots, l_nx_n^k - b_n]^{\mathrm{T}})$ with respect to λ_i obtains

$$\lambda_i = \frac{1}{l_i}. \tag{9}$$

Lemma 1. *When the Mul-GI method with multi-step searches (7)–(9) is applied to the strongly convex quadratic function (5), then in one iteration, the initial point can get to the minimizer $\boldsymbol{x}^1 = \boldsymbol{x}^*$.*

The proof of Lemma 1 is given in the Appendix.

The estimations for the quadratic function $f(x_1, x_2) = x_1^2 + 10x_2^2$ by using the GI algorithm and the Mul-GI algorithm are shown in Table 1.

Remark 1. Lemma 1 and Table 1 imply that a random initial point can arrive at the minimizer $\boldsymbol{x}^*$ in only one iteration by using the Mul-GI method for the quadratic function (5).

3.2. Q is a symmetric and positive definite matrix

In the above subsection, the matrix $\boldsymbol{Q}$ is assumed to be a diagonal and positive definite matrix. When the matrix $\boldsymbol{Q}$ is a symmetric