Journal of Informatics and Mathematical Sciences
Vol. 10, No. 4, pp. 697–702, 2018
ISSN 0975-5748 (online); 0974-875X (print)
Published by RGN Publications
DOI: 10.26713/jims.v10i4.1064

http://www.rgnpublications.com

Research Article

Relationship Between the Fixed Point Theorem and the EM Algorithm

Ahsene Lanani

Department of Mathematics, Faculty of Exact Sciences, University of Freres Mentouri Constantine 1, 25000, Algeria
alanani1@yahoo.fr

Abstract. When we are confronted with solving nonlinear equations which do not admit explicit solutions, we must use approximate methods based on iterative processes or algorithms. One of the best known iterative methods is the fixed point theorem, often applied in analysis or algebra. In our case, we will apply this method in a stochastic context. By means of this application, we show the relationship between this method and the EM algorithm, which is an iterative process, often applied in statistics.

Keywords. EM algorithm; Fixed point; Linear model; Nonlinear equation

MSC. 54H25; 47H10; 55M20

Received: August 21, 2018 **Accepted:** December 27, 2018

This article was submitted to "*International Conference on Mathematics*" organized by Fatih Sultan Mehmet Vakıf University, Topkapı (Cevizlibağ) Campus, Istanbul, Turkey during July 3-6, 2018.

Academic Editors: Dr. Kenan Yildirim, Mus Alparslan University, Turkey
Dr. Gumrah Uysal, Karabuk University, Turkey
Dr. Reza Abazari, University of Tabriz, Iran

1. Introduction

In our research, we are often confronted with solving algebraic equations. When these equations are nonlinear, they often do not admit explicit solutions. So, we try to find approximate solutions by applying numerical methods. These solutions are based on iterative processes or algorithms.

We can quote the dichotomy method; the Newton method and the fixed point theorem; etc. Due to its importance, the latter will be considered. Because of the big number of variants of the fixed point theorem, the knowledge of the existence of fixed points of a function has relevant applications in many branches of analysis; algebra and topology. The theory is well developed by many authors [1,3,8,9]. In this paper, the main objective is not the analytical or algebraic side of this theorem but rather its application in a stochastic context little known compared to the previous ones. Section **??** of this paper consists of recalling the fixed point theorem and the EM algorithm [2,11]. Finally, we show the relationship existing between the two methods. Section **??** is devoted to the application of the EM algorithm in order to estimate the parameters of a linear mixed model. Eventually, section **??** constitutes the concluding part.

2. Fixed Point Theorem and EM Algorithm

2.1 Fixed point theorem

We are interested in the approximate determination of the roots of an equation of the form $f(x) = 0$, where the function f is a numerical function of a real variable defined and continuous in a certain interval $[a,b]$. A fixed point of a function g is a value of x which remains invariant for this function; i.e. any solution of the equation $g(x) = x$ is a fixed point of the g function. The principle of this method is to transform the equation $f(x) = 0$ into an equivalent equation $g(x) = x$ where g is an auxiliary function well chosen. The process is completed by constructing a sequence which converges to the fixed point of g. Let us consider α a fixed point of g, we will determine the elements of this sequence. If x_0 is an initial approximation of α, x_1 is calculated by: $x_1 = g(x_0)$ then x_2 by $x_2 = g(x_1)$ and so on until x_n. This method allows us to construct a recursive sequence $(x_n), n \in N$, defined by: x_0 is an initial approximation of α; $(\forall\ n = 0)$, $x_{n+1} = g(x_n)$. If this sequence converges, its limit will be the solution of the problem. We will not recall the convergence of this sequence which is related to certain conditions on the function g (e.g. g satisfies a Lipschitz condition or is a contraction), because we are more interested in the stochastic side than the analytical one of the application of this method.

2.2 EM Algorithm

Let X be a random variable of density $f(x/\theta)$ where θ is an unknown parameter. Let us suppose that X is not completely observed; i.e. one observes a part Y of X. Let $Y = Y(X)$ be a random variable of density $g(y/\theta)$. Let $t(x)$ be a vector of sufficient (exhaustive statistics for θ). The purpose of the EM algorithm (E for expectation and M for maximization) is to find the value of θ which maximizes the likelihood $g(y/\theta)$ given a value of y as described by [2]. Thereafter, many works concerning its application were carried out by many authors (e.g. [6,11]. This method and especially the convergence of the algorithm is performed by many authors [4,5,7,12,13]. A study related to the mathematical properties of the EM algorithm is carried out in [10].

The likelihood maximization (normal equations) gives the following equation:

$$E(t(x)/\theta) = E(t(x)/y,\theta). \tag{2.1}$$

The EM algorithm uses two stages to solve this θ-equation

E-step: One calculates the quantity:

$$t(x) = E(t(x)/y,\theta). \tag{2.2}$$

M-step: One solves the θ-equation:

$$E(t(x)/\theta) = t(x). \tag{2.3}$$

In other words, in E-step, given an initial value for θ; $\theta^{(p)}$ its value at the stage (p); at the stage $(p+1)$, one calculates the value of $t(x)$ noted by $t^{(p)}$ and which is given by $t^{(p)} = E(t(x)/y,\theta^{(p)})$. In the M-step, given the value of $t^{(p)}$ calculated in the E-step, one solves the equation in $\theta^{(p+1)}$ which is given by:

$$E(t(x)/y,\theta^{(p+1)}) = t^{(p)}. \tag{2.4}$$

2.3 Fixed point method and EM algorithm

Suppose that we want to solve the equation in θ; $F(\theta) = G(\theta)$, where F is some easily invertible function; a simple algorithm is: given $\theta^{(p)}$, the value of the unknown parameter θ obtained in step p; in step $p+1$, we perform two actions:

- Direct calculation of G:

$$t^{(p)} = G(\theta^{(p)}). \tag{2.5}$$

-Inversion of F:

$$\theta^{(p+1)} = F^{-1}(t^{(p)}). \tag{2.6}$$

This algorithm is nothing more than a variant of an algorithm resulting from the fixed point theorem, where F is simply the identity function. Consequently, the two actions become:

- $t^{(p)} = G(\theta^{(p)})$
- $\theta^{(p+1)} = (t^{(p)})$

i.e. $\theta^{(p+1)} = G(\theta^{(p)})$

Under certain conditions (e.g. G satisfies the Lipschitz condition), this algorithm converges to the solution. The aim of this work is not to deal with this aspect of the question but the stochastic context application. The EM algorithm is nothing more than the statistical version of this simple deterministic algorithm. It is sufficient in what precedes to replace $F(\theta)$ and $G(\theta)$ by $E(t(x)/y,\theta)$ and $E(t(x)/\theta)$ respectively. Equations (2.5) and (2.6) are just steps E (direct calculation of G, or calculation of the conditional expectation) and M (easy inversion of F or easy maximum likelihood of f).

3. Application of the EM Algorithm

3.1 Model

Let us consider the following linear mixed model

$$y_i = X_i\alpha + Z_i b_i + e_i, \quad i = 1,\dots m \tag{3.1}$$

where y_i: vector of the responses of dimension $(n_i \times p)$.
X_i: a known $(n_i \times p)$ design matrix linking α to y_i.
α: a $(p \times 1)$ vector of unknown parameters, it is a vector of fixed effects.
Z_i: a known $(n_i \times \mathrm{k})$ design matrix linking b_i to y_i.
b_i: a $(k \times 1)$ vector of unknown parameters, it is a vector of random effects.
b_i is distributed as $N(0,D)$, (normal with mean 0 and covariance matrix D).
$D = D(\theta))$, θ: an unknown $(q \times 1)$ vector.
e_i: vector of the errors which are supposed to be independent and follows $N(0,\sigma^2 I)$; I the identity matrix.

Therefore, the variance-covariance matrix of y_i noted by V_i is given by

$$V_i = Z_i D {Z_i}^t + \sigma^2 I .$$

We want to find the maximum likelihood estimator of θ, parameter generating this matrix of variance-covariance; for that, we apply the EM algorithm. We consider that b_i and e_i are observations in addition to y_i. The sufficient statistics noted by t_1 and t_2, used to estimate θ are $\sum {e_i}^t e_i$, and $\sum b_i {b_i}^t$, respectively.
E-step: we calculate $t_1^{(p)}$ and $t_2^{(p)}$ given by:

$$t_1^{(p)} = E\left(\sum {e_i}^t e_i / y_i,\ \alpha(\theta^{(p)}), \theta^{(p)}\right), \tag{3.2}$$

$$t_2^{(p)} = E\left(\frac{\sum b_i {b_i}^t}{y_i},\ \alpha(\theta^{(p)}), \theta^{(p)}\right) \tag{3.3}$$

M-step: we solve the equations in $\theta^{(p+1)}$

$$E\left(\sum {e_i}^t e_i,\ \alpha\left(\theta^{(p+1)}\right), \theta^{(p+1)}\right) = t_1^{(p)} \tag{3.4}$$

$$E\left(\sum b_i {b_i}^t / \alpha\left(\theta^{(p+1)}\right), \theta^{(p+1)}\right) = t_2^{(p)} \tag{3.5}$$

3.2 Calculus

The parameter θ is composed of σ^2, parameter generating t_1 and of $\left(\frac{1}{2}\right) k(k+1)$ component of t_2.
M-Step: in this stage, we will use the expressions (3.6) and (3.7) below

$$\sigma^2 = \frac{\sum_1^m {e_i}^t e_i}{\sum_1^m n_i} = \frac{t_1}{\sum_1^m n_i} \tag{3.6}$$

$$D = m^{-1} \sum_1^m b_i {b_i}^t = \frac{t_2}{m} \tag{3.7}$$

E-Step: having a preliminary value for θ (initial value), we then calculate the estimators of statistics t_1 and t_2.

Therefore, at the (p) stage or in $\theta^{(p)}$ (preliminary value), we have the expressions $t_1^{(p)}$ and $t_2^{(p)}$ which are given by the formulas (3.8) and (3.9) below

$$\begin{aligned} t_1^{(p)} &= E\left(\sum {e_i}^t e_i / y_i,\ \alpha(\theta^{(p)}), \theta^{(p)}\right) \\ &= \sum_1^m e_i(\theta^{(p)})^t e_i(\theta^{(p)}) + \operatorname{tr}\ \operatorname{var}(e_i / y_i,\ \alpha(\theta^{(p)}), \theta^{(p)}) \end{aligned} \tag{3.8}$$

where *tr* and *var* means trace and variance respectively.

$$t_2^{(p)} = E\left(\sum b_i b_i{}^t/y_i,\ \alpha(\theta^{(p)}),\theta^{(p)}\right)$$
$$= \sum_1^m b_i(\theta^{(p)})b_i(\theta^{(p)})^t + \mathrm{var}(b_i/y_i,\ \alpha(\theta^{(p)}),\theta^{(p)}) \tag{3.9}$$

where

$$e_i(\theta^{(p)}) = E(e_i/y_i, \alpha(\theta^{(p)}),\theta^{(p)}) = y_i - X_i\ \alpha(\theta^{(p)}) - Z_i\ b_i(\theta^{(p)})$$

with $\alpha(\theta^{(p)})$ and $b_i(\theta^{(p)})$ given by

$$\alpha(\theta^{(p)}) = \left(\sum_1^m X_i{}^t W_i(\theta^{(p)})X_i\right)^{-1}\left(\sum_1^m X_i{}^t W_i(\theta^{(p)})y_i\right)$$
$$b_i(\theta^{(p)}) = D(\theta^{(p)})Z_i{}^t W_i(\theta^{(p)})(y_i - X_i\ \alpha(\theta^{(p)})$$

where $W_i = V_i{}^{-1}$.

Recall that the estimators of α and b_i are, the maximum likelihood estimator for α and the estimator of generalized least squares or the empirical Bayes estimator for b_i which is given by:

$$b_i(\theta^{(p)}) = E(b_i/y_i, \alpha(\theta^{(p)}),\theta^{(p)})$$

To calculate $E(e_i/y_i, \alpha(\theta^{(p)}),\theta^{(p)})$ and $\mathrm{tr}\,\mathrm{var}(e_i/y_i, \alpha(\theta^{(p)}),\theta^{(p)})$, it is necessary to calculate the distribution of e_i conditionally at $(y_i, \alpha(\theta^{(p)}),\theta^{(p)})$. The same is done to calculate $b_i(\theta^{(p)})$ and $\mathrm{var}(b_i/y_i, \alpha(\theta^{(p)}),\theta^{(p)})$.

To have the maximum likelihood estimator of θ, that we will note by θ_M, we start with a suitable initial value of θ, we make then iterations between (3.8) and (3.9) stage defining the E-step, and (3.6) and (3.7) stage defining the M-step. At convergence, we do not have only θ_M, but also $\alpha(\theta_M)$ and $b_i(\theta_M)$ of the calculation of the last E-step.

4. Conclusion

We are interested in fixed point theorem; firstly because its application in the stochastic context is little known and secondly the hidden side of the passage of the fixed point theorem to the EM algorithm is not obvious. In this paper, we demonstrated the relationship between the two methods. Finally, we emphasized the importance of the EM algorithm to estimate the parameters of a linear mixed model.

Competing Interests

The author declares that he has no competing interests.

Authors' Contributions

The author wrote, read and approved the final manuscript.

References

[1] R. P. Agarwal, M. Meehan and D. O'Regan, *Fixed Point Theory and Applications*, Cambridge University Press (2001).

[2] A. P. Dempster, N. M. Laird and D. B. Rubin, Maximum likelihood with incomplete data via the E-M algorithm, *J. Roy. Stat. Soc.* **B39** (1977), 1 – 38.

[3] J. Dugundji and A. Granas, *Fixed Point Theory*, Springer-Verlag (2003).

[4] V. Florin, Parameter convergence for EM and MM algorithms, *Statistica Sinica* **15** (2005), 831 – 840.

[5] J. L. Foulley and D. A. Van Dyk, The PX EM algorithm for fast fitting of Henderson's mixed model, *Genetics Selection Evolution* **32** (2000), 143 – 163.

[6] J. L. Foulley, F. Jaffrézic and C. R. Granié, EM-REML estimation of covariance parameters in gaussian mixed models for longitudinal data analysis, *Genetics Selection Evolution* **32** (2000), 129 – 141.

[7] F. W. Homer and N. Peng, Anderson acceleration for fixed-point iterations, *SIAM J. Num. Anal.* **49**(4) (2011), 1715 – 1735.

[8] W. A. Kirk and M. A. Khamsi, *An Introduction to Metric Spaces and Fixed Point Theory*, John Wiley, New York (2001).

[9] W. A. Kirk and S. Brailey, *Handbook of Metric Fixed Point Theory*, Springer-Verlag (2001).

[10] K. Kubjas, E. Robeva and B. Sturmfels, Fixed points of the EM algorithm and nonnegative rank boundaries, *The Annals of Stat.* **43**(1) (2015), 422 – 461.

[11] N. M. Laird and J. H. Ware, Random effect models for longitudinal data, *Biometrics* **38** (1982), 963 – 974.

[12] N. M. Laird, N. Lange and D. Stram, Maximum likelihood computations with repeated measures: application of The EM algorithm, *J. Amer. Statist. Assoc.* **82**(397) (1987), 97 – 105.

[13] M. J. Lindstrom and D. M. Bates, Newton-Raphson and EM algorithm for linear mixed effect models for repeated-measures data, *J. Amer. Statist. Assoc.* **83**(404) (1988), 1014 – 1022.