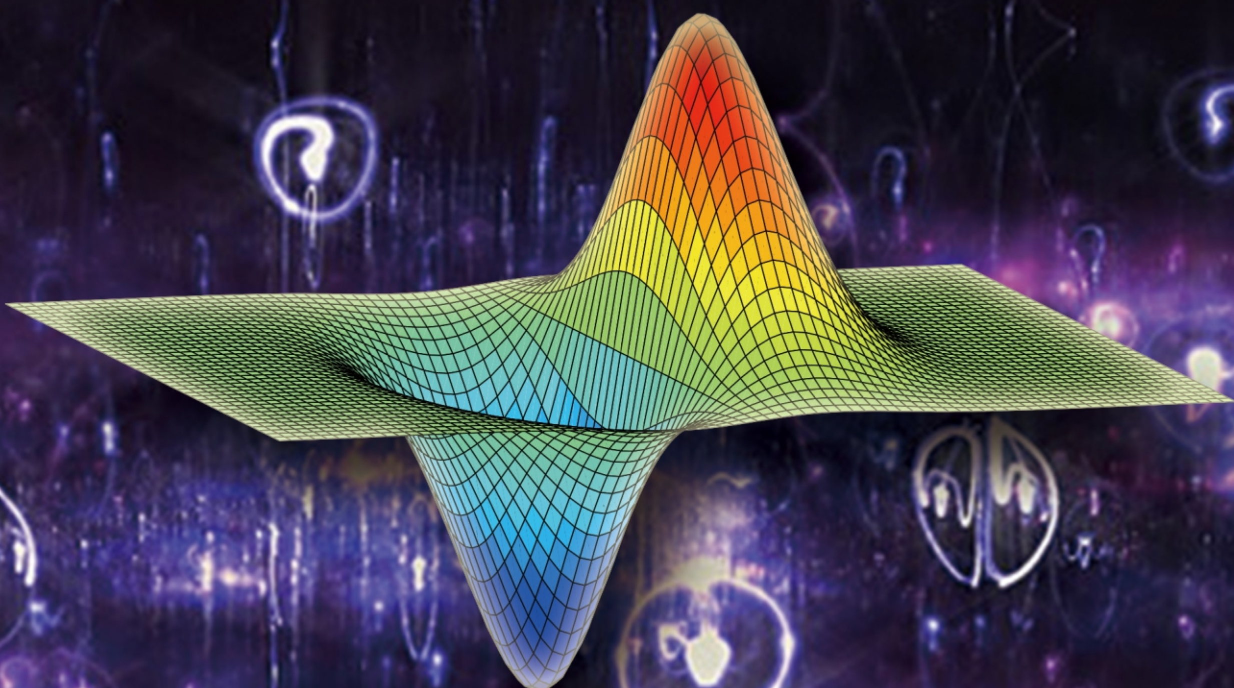
ISSN: 2160-0368 Volume 10, Number 3, March 2020

Scientific Research Publishing

Advances in Pure Mathematics



ISSN: 2160-0368

https://www.scirp.org/journal/apm

Journal Editorial Board

ISSN 2160-0368 (Print) ISSN 2160-0384 (Online)
https://www.scirp.org/journal/apm

Scientific Research Publishing

Advances in Pure Mathematics, 2020, 10, 101-154
https://www.scirp.org/journal/apm
ISSN Online: 2160-0384
ISSN Print: 2160-0368

Table of Contents

Volume 10 Number 3 March 2020

Advances in Pure Mathematics (APM)
Journal Information

SUBSCRIPTIONS

The *Advances in Pure Mathematics* (Online at Scientific Research Publishing, https://www.scirp.org/) is published monthly by Scientific Research Publishing, Inc., USA.

Subscription rates:
Print: $79 per issue.
To subscribe, please contact Journals Subscriptions Department, E-mail: sub@scirp.org

SERVICES

Advertisements
Advertisement Sales Department, E-mail: service@scirp.org

Reprints (minimum quantity 100 copies)
Reprints Co-ordinator, Scientific Research Publishing, Inc., USA.
E-mail: sub@scirp.org

COPYRIGHT

PRODUCTION INFORMATION

For manuscripts that have been accepted for publication, please contact:
E-mail: apm@scirp.org

Advances in Pure Mathematics, 2020, 10, 101-113
https://www.scirp.org/journal/apm
ISSN Online: 2160-0384
ISSN Print: 2160-0368

Variable Selection in Finite Mixture of Time-Varying Regression Models

Jing Liu, Wanzhou Ye

College of Science, Shanghai University, Shanghai, China
Email: wzhy@shu.edu.cn

How to cite this paper: Liu, J. and Ye, W.Z. (2020) Variable Selection in Finite Mixture of Time-Varying Regression Models. *Advances in Pure Mathematics*, **10**, 101-113.
https://doi.org/10.4236/apm.2020.103007

Received: January 30, 2020
Accepted: March 3, 2020
Published: March 6, 2020

Abstract

In this paper, we research the regression problem of time series data from heterogeneous populations on the basis of the finite mixture regression model. We propose two finite mixed time-varying regression models to solve this. A regularization method for variable selection of the models is proposed, which is a mixture of the appropriate penalty functions and l_2 penalty. A Block-wise minimization maximization (MM) algorithm is used for maximum penalized log quasi-likelihood estimation of these models. The procedure is illustrated by analyzing simulations and with an application to analyze the behavior of urban vehicular traffic of the city of São Paulo in the period from 14 to 18 December 2009, which shows that the proposed models outperform the FMR models.

Keywords

Mixture Regression Models, GARCH, Block-Wise MM algorithm, LASSO, SCAD

1. Introduction

The problem of variable selection in FMR models has been widely discussed [1] [2] [3]. When a response variable $\boldsymbol{y}$ with a finite mixture distribution depends on covariates $\boldsymbol{x}$, we obtain a finite mixture of regression (FMR) model. The FMR model with K components can be given as follows [3]:

$$f(y;\boldsymbol{x},\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k f(y;\eta_k(\boldsymbol{x}),\phi_k) \tag{1}$$

where $\boldsymbol{y}$ is an independent and identically distributed (IID) response and $\boldsymbol{x}$ is a $p\times 1$ vector of covariates. $\boldsymbol{\pi} = (\pi_1,\cdots,\pi_k)^{\mathrm{T}}$ denotes the mixing proportions satisfying $0<\pi_k<1$, $\sum_{k=1}^{K}\pi_k = 1$. $f(y;\eta_k(\boldsymbol{x}),\phi_k)$ is the kth mixture component density. $\eta_k(\boldsymbol{x}) = h(\alpha_k + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_k)$ for $k = 1,\cdots,K$, for a given link function $h(\cdot)$, and a dispersion parameter ϕ_k.

However, in some situations, observations were not independent. As pointed out in [2], in the analysis of the PD data, observations from each patient over time were assumed to be independent to facilitate the analysis and comparison with results from the literature. However, the validity of such assumption may be questionable. Whereupon, we consider a situation that observations were time series.

The generalised autoregressive conditional heteroskedasticity (GARCH) model is widely used in time series analysis. A mixture generalized autoregressive conditional heteroscedastic (MGARCH) model was pointed out in [4]. [5] generalized the MixN-GARCH model by relaxing the assumption of constant mixing weights. Whereupon, we combine the GARCH model and the FMR model to discuss the above problem.

There has been extensive studies about variable selection methods. A recent review of the literature regarding the variable selection problem in FMR models can be found in [6]. There are a general family of penalty functions, including the least absolute shrinkage and selection operator (LASSO), the minimax concave penalty (MCP) and the smoothly clipped absolute deviation (SCAD) in [2] and [7].

The method of the maximum penalized log-likelihood (MPL) estimation is usually the EM algorithm. [8] proposed a new algorithm (block-wise MM) for the MPL estimation of the L-MLR model. It was proved to have some desirable features such as coordinate-wise updates of parameters, monotonicity of the penalized likelihood sequence, and global convergence of the estimates to a stationary point of the penalized loglikelihood function, which are missing in the commonly used approximate-EM algorithm presented in [3].

The rest of the paper is organized as follows: in Section 2, the definition of finite mixture of time-varying regression Models and in Section 3, feature selection methods are discussed. In Section 4, the block-wise MM algorithm for its estimation and the BIC for choosing tuning parameters and components are presented, and the example of the Gaussian distribution is derived. Simulation studies on the performance of the new variable selection methods are then provided in Section 5. In Section 6, analysis of a real data set illustrates the use of the procedure. Finally, conclusions are given in Section 7.

2. Finite Mixture of Time-Varying Regression Models

2.1. Finite Mixture of Autoregression Models

Let $\{y_t; t = 1,\cdots,n\}$ be a response variable which is a time series. $\{\boldsymbol{x}_t; t = 1,\cdots,n\}$ is a p-dimensional vector of covariates, and each of them is a time series. For an FM-AR(d) model with K components, the conditional density function for observation t is given as follows:

$$f(y_t; \boldsymbol{x}_t, \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k f(y_t; \eta_k(\boldsymbol{x}_t), \phi_k), \tag{2}$$

where

$$\eta_k(\mathbf{x}_t) = h\left(\alpha_k + \mathbf{x}_t^{\mathrm{T}}\boldsymbol{\beta}_{k1} + \mathbf{x}_{t-1}^{\mathrm{T}}\boldsymbol{\beta}_{k2} + \cdots + \mathbf{x}_{t-d}^{\mathrm{T}}\boldsymbol{\beta}_{kd}\right), \tag{3}$$

for $k = 1,\cdots,K$, for a given link function $h(\cdot)$, and a dispersion parameter ϕ_{kt}.

The master vector of all parameters is given by $\boldsymbol{\theta} = \left(\boldsymbol{\pi}^{\mathrm{T}}, \boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\phi}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}}\right)^{\mathrm{T}}$, with

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_{11} & \cdots & \boldsymbol{\beta}_{1d} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\beta}_{K1} & \cdots & \boldsymbol{\beta}_{Kd} \end{pmatrix}, \tag{4}$$

where $\boldsymbol{\beta}_{ki} = \left(\beta_{ki1},\cdots,\beta_{kip}\right)^{\mathrm{T}} \in \mathbb{R}^p$, $i = 1,\cdots,d$. Let $\tilde{x}_t = \left(\mathbf{x}_t^{\mathrm{T}}, \mathbf{x}_{t-1}^{\mathrm{T}},\cdots,\mathbf{x}_{t-d}^{\mathrm{T}}\right)$, and $\tilde{\boldsymbol{\beta}} = \left(\boldsymbol{\beta}_{k1},\cdots,\boldsymbol{\beta}_{kd}\right)^{\mathrm{T}}$, (3) can be rewrote as $\eta_k(\mathbf{x}_t) = h\left(\alpha_k + \tilde{x}_t\tilde{\boldsymbol{\beta}}\right)$.

2.2. Finite Mixture of GARCH Models

Let $\{y_t; t = 1,\cdots,n\}$ be a response variable which is a time series. Let $\{\mathbf{x}_t; t = 1,\cdots,n\}$ is a *p*-dimensional vector of covariates, and each of them is a time series. For some distributions with unequal dispersion parameter ϕ_k, we propose the FM-GARCH models. For an FM-GARCH (*d*,*M*,*S*) model with *K* components, the conditional density function for observation *t* is given as follows:

$$f(y_t; \mathbf{x}_t, \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k f\left(y_t; \eta_k(\mathbf{x}_t), \phi_{kt}\right), \tag{5}$$

where $\eta_k(\mathbf{x}_t) = h\left(\alpha_k + \tilde{x}_t\tilde{\boldsymbol{\beta}}\right)$ for $k = 1,\cdots,K$, for a given link function $h(\cdot)$, and a conditional heteroscedastic (a dispersion parameter)

$$\phi_{kt} = \gamma_{0k} + \sum_{m=1}^{M} \gamma_{km}\epsilon_{k,t-m} + \sum_{s=1}^{S} \delta_{ks}\phi_{k,t-s}, \tag{6}$$

where $\gamma_{0k} > 0$, $\gamma_{km} \geq 0$, $\delta_{ks} \geq 0$, and $\epsilon_{kt} = \phi_{kt}e_{kt}$, e_{kt} is an independent and identically distributed series with mean zero and variance unity.

The master vector of all parameters is given by $\boldsymbol{\theta} = \left(\boldsymbol{\pi}^{\mathrm{T}}, \boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\gamma}_0^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\gamma}^{\mathrm{T}}, \boldsymbol{\delta}^{\mathrm{T}}\right)^{\mathrm{T}}$, with $\boldsymbol{\gamma}_0 = \left(\gamma_{01},\cdots,\gamma_{0K}\right)^{\mathrm{T}}$, $\boldsymbol{\gamma} = \left(\boldsymbol{\gamma}_1,\cdots,\boldsymbol{\gamma}_K\right)^{\mathrm{T}}$, $\boldsymbol{\gamma}_k = \left(\gamma_{k1},\gamma_{k2},\cdots,\gamma_{kM}\right)^{\mathrm{T}}$, and $\boldsymbol{\delta} = \left(\boldsymbol{\delta}_1,\cdots,\boldsymbol{\delta}_K\right)^{\mathrm{T}}$, $\boldsymbol{\delta}_k = \left(\delta_{k1},\delta_{k2},\cdots,\delta_{kS}\right)^{\mathrm{T}}$.

3. Feature Selection Method

Let $\{(\mathbf{x}_t, y_t); t = 1,\cdots,n\}$ be a sample of observations from the FM-AR or FM-GARCH model. The quasi-likelihood function of the parameter $\boldsymbol{\theta}$ is given by [9]

$$\mathrm{L}_n(\boldsymbol{\theta}) = \prod_{t=1}^{n} f(y_t; \mathbf{x}_t, \boldsymbol{\theta}) = \prod_{t=1}^{n}\left\{\sum_{k=1}^{K} \pi_k f\left(y_t; \eta_k(\mathbf{x}_t), \phi_{kt}\right)\right\}. \tag{7}$$

The log quasi-likelihood function of the parameter $\boldsymbol{\theta}$ is given by

$$\mathcal{L}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \log \sum_{k=1}^{K} \pi_k f\left(y_t; \eta_k(\mathbf{x}_t), \phi_{kt}\right). \tag{8}$$

When the effect of a component of $\mathbf{x}$ is not significant, the corresponding ordinary maximum quasi-likelihood estimate is often close to 0, but not equal to 0. Thus this covariate is not excluded from the model. Inspired by an idea of [2], we estimate $\boldsymbol{\theta}$ by maximizing the penalized log quasi-likelihood function (MPLQ) for the model

$$\mathcal{F}_n(\boldsymbol{\theta}) = \mathcal{L}_n(\boldsymbol{\theta}) - \mathcal{P}_n(\boldsymbol{\theta}), \tag{9}$$

with the mixture penalty (or regularization) function:

$$\mathcal{P}_{nk}(\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \sum_{i=1}^{d}\sum_{j=1}^{p} p_n(\beta_{kij}; \lambda_{nk}) + \frac{1}{2}\sum_{k=1}^{K} \pi_k \sum_{i=1}^{d}\sum_{j=1}^{p} \upsilon_{nk}\beta_{kij}^2, \tag{10}$$

for some ridge tuning parameter $\upsilon_{nk} \ge 0$, and $p_n(\beta_{kij}; \lambda_{nk})$ is a nonnegative penalty function. In the penalty function $\mathcal{P}_n(\boldsymbol{\theta})$, the amount of l_2 penalty imposed on the componentwise regression coefficients β_{kij}'s are chosen proportional to π_k. The functions $p_n(\beta_{kij}; \lambda_{nk})$ are designed to identify the no significant coefficients β_{kij}'s in the mixture components $f(y_t; \eta_i(\mathbf{x}_t), \phi_{kt})$. General regularity conditions about the $p_n(\beta_{kij}; \lambda_{nk})$ is given in [2] [3].

We estimate the new method using the following well-known penalty (or regularization) functions:

- LASSO penalty: $p_n(\beta; \lambda_{nk}) = \lambda_{nk}|\beta|$.
- MCP penalty: $p_n'(\beta; \lambda_{nk}) = (\lambda_{nk} - nb_{nk}|\beta|)_+$.
- SCAD penalty:

$$p_n'(\beta; \lambda_{nk}) = \lambda_{nk} I(n|\beta| < \lambda_{nk}) + \frac{(a_{nk}\lambda_{nk} - n|\beta|)_+}{a_{nk} - 1} I(n|\beta| > \lambda_{nk}).$$

Here, I is the indicative function. The constant $a_{nk} \ge 2$ and $b_{nk} \ge 0$ pointed in [2], and LASSO tuning parameter $\lambda_{nk} \ge 0$, which controls the amount of penalty. The asymptotic properties about these penalty functions can be analogously derived in [3] and [2]. We call the penalty function $\mathcal{P}_{nk}(\boldsymbol{\theta})$ in (10) constructed from LASSO, MCP, SCAD jointly with the mixed L_2-norm as MIXLASSO-ML$_2$, MIXMCP-ML$_2$, MIXSCAD-ML$_2$ penalties.

4. Numerical Solutions

A new method for maximizing the penalized log-likelihood function is the block-wise Minorization Maximization (MM) algorithm inspired by [8], which is also known as block successive lower-bound maximization (BSLM) algorithm in the language of [10]. At each iteration of the method, the function is maximized with respect to a single block of variables while the rest of the blocks are held fixed. We shall now proceed to describe the general framework of the algorithm.

4.1. Maximization of the Penalized Log-Likelihood Function

We follow the approach of [8] and minorize the ε-approximate of -$\mathcal{P}_n(\boldsymbol{\theta})$ by

$$\mathrm{G}_1(\boldsymbol{\theta}; \boldsymbol{\theta}^{(r)}) = -\frac{1}{2}\sum_{k=1}^{K} \pi_i \sum_{j=1}^{d}\sum_{k=1}^{p} p_n\left(\frac{\beta_{ijk}^2}{w_{ijk}^{(r)}}; \lambda_{ni}\right) - \frac{1}{2}\sum_{k=1}^{K} \pi_i \sum_{j=1}^{d}\sum_{k=1}^{p} \upsilon_{ni}\beta_{ijk}^2 + C_1(\boldsymbol{\theta}^{(r)}), \tag{11}$$

where $w_{ijk}^{(r)} = \sqrt{\beta_{ijk}^{2(r)} + \varepsilon^2}$, for some $\varepsilon > 0$, and

$$C_1\left(\boldsymbol{\theta}^{(r)}\right) = -\frac{\varepsilon^2}{2}\sum_{k=1}^{K}\pi_i\sum_{j=1}^{d}\sum_{k=1}^{p}p_n^{-1}\left(w_{ijk}^{(r)};\lambda_{ni}\right) - \frac{1}{2}\sum_{k=1}^{K}\pi_i\sum_{j=1}^{d}\sum_{k=1}^{p}p_n\left(w_{ijk}^{(r)};\lambda_{ni}\right). \tag{12}$$

Moreover, minorize the log quasi-likelihood function $\mathcal{L}_n(\boldsymbol{\theta})$ by

$$\begin{aligned} G_2\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}\right) &= \sum_{k=1}^{K}\sum_{t=1}^{n}\tau_{kt}^{(r)}\log\pi_i + \sum_{k=1}^{K}\sum_{t=1}^{n}\tau_{kt}^{(r)}\log f\left(y_t;\eta_i\left(x_t\right),\phi_{kt}\right) \\ &\quad - \sum_{k=1}^{K}\sum_{t=1}^{n}\tau_{kt}^{(r)}\log\tau_{kt}^{(r)}, \end{aligned} \tag{13}$$

where $\tau_{kt}^{(r)} = \pi_i^{(r)} f\left(y_t;\eta_i^{(r)}\left(\mathbf{x}_t\right),\phi_{kt}^{(r)}\right) \Big/ f\left(y_t;\mathbf{x}_t,\boldsymbol{\theta}^{(r)}\right)$.

Note that $\tau_{kt}^{(r)}$ and $G_2\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}\right)$ are analogous to the posterior probability and the expected complete-data log-likelihood function of the expectation-maximization algorithm respectively.

The block-wise MM algorithm maximizes $\mathcal{F}_n(\boldsymbol{\theta})$ iteratively in the following two steps:

- Block-wise Minorization-step. Conditioned on the *r*th iterate $\boldsymbol{\theta}^{(r)}$, the FM-GARCH model can be block-wise minorized in the coordinates of the parameter components $\boldsymbol{\pi}$, $\boldsymbol{\alpha}$, γ_0, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, and $\boldsymbol{\beta}$, via the minorizers

$$G_{\pi}(\boldsymbol{\pi};\boldsymbol{\theta}) = G_2\left(\boldsymbol{\pi},\boldsymbol{\alpha}^{(r)},\gamma_0^{(r)},\boldsymbol{\beta}^{(r)},\boldsymbol{\gamma}^{(r)},\boldsymbol{\delta}^{(r)};\boldsymbol{\theta}^{(r)}\right) - \mathcal{P}_n\left(\boldsymbol{\pi},\boldsymbol{\beta}^{(r)}\right), \tag{14}$$

$$G_{\alpha,\gamma_0}\left(\boldsymbol{\alpha},\gamma_0;\boldsymbol{\theta}^{(r)}\right) = G_2\left(\boldsymbol{\pi}^{(r)},\boldsymbol{\alpha},\gamma_0,\boldsymbol{\beta}^{(r)},\boldsymbol{\gamma}^{(r)},\boldsymbol{\delta}^{(r)};\boldsymbol{\theta}^{(r)}\right) - \mathcal{P}_n\left(\boldsymbol{\theta}^{(r)}\right), \tag{15}$$

$$G_{\gamma,\delta}\left(\boldsymbol{\gamma},\boldsymbol{\delta};\boldsymbol{\theta}^{(r)}\right) = G_2\left(\boldsymbol{\pi}^{(r)},\boldsymbol{\alpha}^{(r)},\gamma_0^{(r)},\boldsymbol{\beta}^{(r)},\boldsymbol{\gamma},\boldsymbol{\delta};\boldsymbol{\theta}^{(r)}\right) - \mathcal{P}_n\left(\boldsymbol{\theta}^{(r)}\right), \tag{16}$$

$$G_{\beta}\left(\boldsymbol{\beta};\boldsymbol{\theta}^{(r)}\right) = G_1\left(\boldsymbol{\pi}^{(r)},\boldsymbol{\beta};\boldsymbol{\theta}^{(r)}\right) + G_2\left(\boldsymbol{\pi}^{(r)},\boldsymbol{\alpha}^{(r)},\gamma_0^{(r)},\boldsymbol{\beta},\boldsymbol{\gamma}^{(r)},\boldsymbol{\delta}^{(r)};\boldsymbol{\theta}^{(r)}\right), \tag{17}$$

respectively. Similar block-wise minorized can be made for FM-AR model.

- Block-wise Maximization-step. Upon finding the appropriate set of block-wise minorizers of $\mathcal{F}_n(\boldsymbol{\theta})$, we can maximize (14) to compute the $(r+1)$ th iterate block-wise update of $\boldsymbol{\pi}$. Solving for the appropriate root of the FOC (first-order condition) for the Lagrangian, we can compute the $(r+1)$ th iterate block-wise update

$$\pi_k^{(r+1)} = \frac{\sum_{t=1}^{n}\tau_{kt}^{(r)}}{\zeta^* + z_k}, \tag{18}$$

for each *k*, where $z_k = \sum_{i=1}^{d}\sum_{j=1}^{p}p_n\left(\beta_{kij};\lambda_{ni}\right) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{p}\upsilon_{ni}\beta_{kij}^2$, and ζ^* is the unique root of

$$\sum_{k=1}^{K}\frac{\sum_{t=1}^{n}\tau_{kt}^{(r)}}{\zeta^* + z_k} - 1 = 0, \tag{19}$$

in the interval $\left(z^*,\infty\right)$, and $z^* = -\min_{k=1,\cdots,K}\left\{z_k\right\}$.

The block-wise updates for $\boldsymbol{\alpha}$, γ_0, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, and $\boldsymbol{\beta}$ can be obtained by solving (15)-(17) via the first-order condition equal to 0.

We now present a example of the Gaussian FM-GARCH model to specify the procedure described above, and give the following Lemma 1 about a useful minorizer for the MPL estimation of the Gaussian FM-GARCH model, which can be found in [11].

Lemma 1 *if* $\Theta = (0,\infty)$, *then the function* $1\big/\sum_{i=1}^{q} c_i\theta_i$ *satisfy that*

$$\frac{1}{\sum_{i=1}^{q} c_i\theta_i} \le \sum_{i=1}^{q} \frac{c_i\varphi_i^2}{\left(\sum_{i=1}^{q} c_i\varphi_i\right)^2 \theta_i}. \tag{20}$$

Example 1 *We consider the Gaussian FM-GARCH Model,*

$$f(y_t;\boldsymbol{x}_t,\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k N\left(y_t;\eta_k(\boldsymbol{x}_t),\sigma_{kt}^2\right), \tag{21}$$

where $\eta_k(\boldsymbol{x}_t) = h\left(\alpha_k + \tilde{\boldsymbol{x}}_t\tilde{\boldsymbol{\beta}}\right)$, *and* $\sigma_{kt}^2 = \gamma_{0k} + \sum_{m=1}^{M} \gamma_{km}\varepsilon_{k,t-m}^2 + \sum_{s=1}^{S} \delta_{ks}\sigma_{k,t-s}^2$. *Here,* $\varepsilon_{kt} = \sigma_{kt}e_{kt}$, *and* e_{kt} *is an independent and identically distributed series with mean zero and variance unity.*

According to [8], *and using Lemma* 1, *we can obtain the further minorizer of Gaussian FM-GARCH by*

$$\begin{aligned}\tilde{G}_2\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}\right) = &\sum_{k=1}^{K}\sum_{t=1}^{n} \tau_{kt}^{(r)} \log \pi_k - \frac{1}{2}\sum_{k=1}^{K}\sum_{t=1}^{n} \tau_{kt}^{(r)} \log \sigma_{kt}^2 \\ &- \frac{1}{2pd}\sum_{k=1}^{K}\sum_{j=1}^{pd}\sum_{t=1}^{n} \frac{\tau_t^{(r)}}{\sigma_{kt}^2}\left(y_t - \alpha_k - pdx_{tj}\left(\beta_{kj} - \beta_{kj}^{(r)}\right) - \tilde{\boldsymbol{x}}_t^{\mathrm{T}}\tilde{\boldsymbol{\beta}}_k^{(r)}\right)^2 \\ &+ C_2\left(\boldsymbol{\theta}^{(r)}\right),\end{aligned} \tag{22}$$

where $\tau_{kt}^{(r)} = \pi_k^{(r)} N\left(y_t;\alpha_k^{(r)} + \tilde{\boldsymbol{x}}_t^{\mathrm{T}}\tilde{\boldsymbol{\beta}}_k^{(r)},\sigma_{kt}^{2(r)}\right)\Big/\sum_{k=1}^{K}\pi_k N\left(y_t;\alpha_k^{(r)} + \tilde{\boldsymbol{x}}_t^{\mathrm{T}}\tilde{\boldsymbol{\beta}}_k^{(r)},\sigma_{kt}^{2(r)}\right)$, *and*

$$C_2\left(\boldsymbol{\theta}^{(r)}\right) = -\frac{n}{2}\log(2\pi) - \sum_{k=1}^{K}\sum_{t=1}^{n} \tau_{kt}^{(r)} \log \tau_{kt}^{(r)}.$$

The block-wise updates of $\boldsymbol{\pi}$ *from Gaussian FM-GARCH Model come from* (18), *and the block-wise updates for* $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, *and* $\boldsymbol{\delta}$, *can be obtained from* (15)-(16) *via the first-order condition equal to* 0. *By doing so, we obtain the coordinate-wise updates for* $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}_0$ *block*

$$\alpha_k^{(r+1)} = \frac{\sum_{t=1}^{n} \tau_{kt}^{(r)}\left(y_t - \tilde{\boldsymbol{x}}_t^{\mathrm{T}}\tilde{\boldsymbol{\beta}}_k^{(r)}\right)\Big/\sigma_{kt}^{2(r)}}{\sum_{t=1}^{n} \tau_{kt}^{(r)}\Big/\sigma_{kt}^{2(r)}}, \tag{23}$$

$$\gamma_{0k}^{(r+1)} = \frac{\sum_{t=1}^{n} \tau_{kt}^{(r)}\gamma_{0k}^{2(r)}\left(y_t - \alpha_k^{(r+1)} - \tilde{\boldsymbol{x}}_t^{\mathrm{T}}\tilde{\boldsymbol{\beta}}_k^{(r)}\right)^2\Big/\left(\sigma_{kt}^{2(r)}\right)^2}{\sum_{t=1}^{n} \tau_{kt}^{(r)}\gamma_{0k}^{(r)}\Big/\sigma_{kt}^{2(r)}}, \tag{24}$$

for each k. Moreover, the coordinate-wise updates for the $\boldsymbol{\gamma}$ *and* $\boldsymbol{\delta}$ *block*

$$\gamma_{km}^{(r+1)} = \frac{\sum_{t=1}^{n} \tau_{kt}^{(r)}\gamma_{km}^{2(r)}\varepsilon_{k,t-m}^2\left(y_t - \alpha_k^{(r+1)} - \tilde{\boldsymbol{x}}_t^{\mathrm{T}}\tilde{\boldsymbol{\beta}}_k^{(r)}\right)^2\Big/\left(\sigma_{kt}^{2(r)}\right)^2}{\sum_{t=1}^{n} \tau_{kt}^{(r)}\gamma_{km}^{(r)}\varepsilon_{k,t-m}^2\Big/\sigma_{kt}^{2(r)}}, \tag{25}$$

$$\delta_{ks}^{(r+1)} = \frac{\sum_{t=1}^{n} \tau_{kt}^{(r)}\delta_{ks}^{2(r)}\sigma_{k,t-s}^{2(r)}\left(y_t - \alpha_k^{(r+1)} - \tilde{\boldsymbol{x}}_t^{\mathrm{T}}\tilde{\boldsymbol{\beta}}_k^{(r)}\right)^2\Big/\left(\sigma_{kt}^{2(r)}\right)^2}{\sum_{t=1}^{n} \tau_{kt}^{(r)}\delta_{ks}^{(r)}\sigma_{k,t-s}^{2(r)}\Big/\sigma_{kt}^{2(r)}}, \tag{26}$$

for each k, m, and s. Finally, making the substitute (22) *into* (17), *the coordinate-wise updates for the* $\boldsymbol{\beta}$ *block*

$$\beta_{kj}^{(r+1)} = \frac{pd\,\beta_{kj}^{(r)} \sum_{t=1}^{n} \tau_{kt}^{(r)} x_{tj}^{2} \big/ \sigma_{kt}^{2(r)} + \sum_{t=1}^{n} \tau_{kt}^{(r)} x_{tj} \left(y_t - \alpha_k^{(r)} - \tilde{\mathbf{x}}_t^{\mathrm{T}} \tilde{\boldsymbol{\beta}}_k^{(r)} \right) \big/ \sigma_{kt}^{2(r)}}{\mathrm{G}_1'\left(\boldsymbol{\pi}^{(r)}, \boldsymbol{\beta}; \boldsymbol{\theta}^{(r)}\right) \big/ \beta_{kj} + pd \sum_{t=1}^{n} \tau_{kt}^{(r)} x_{tj}^{2} \big/ \sigma_{kt}^{2(r)}}, \quad (27)$$

for each k and $j = 1, \cdots, pd$, *where* $\mathrm{G}_1'\left(\boldsymbol{\pi}^{(r)}, \boldsymbol{\beta}; \boldsymbol{\theta}^{(r)}\right)$ *is the first derivative of* (11) *with respect to* $\boldsymbol{\beta}$.

Note that (15)-(17) *from Gaussian FM-GARCH Model are concave in the alternative parameterization* $\boldsymbol{\pi}$, $\boldsymbol{\alpha}$, γ_0, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, and $\boldsymbol{\beta}$, *thus* (23)-(27) *globally maximize* (15)-(17) *over the parameter space.*

4.2. Selection of Thresholding Parameters and Components

To implement the methods described in Sections 3 and 4.1, we need to select the size of the tuning parameters λ_{nk} and υ_{nk}, the constant a_{nk} and b_{nk}, for $k = 1, \cdots, K$, and components *K*. The current theory provides some guidance on the order of $\boldsymbol{\lambda}$ in [3] and [8] by using generalized cross validation (GCV) and Bayesian Information Criterion (BIC), to ensure the sparsity property. Following the example of [8], we develop a suitable BIC criterion for the FM-AR and FM-GARCH models. Let $\boldsymbol{\Psi} = (\boldsymbol{\lambda}, \boldsymbol{\upsilon}, \boldsymbol{a}, \boldsymbol{b}, K)$, and they are chosen one at a time by minimizing

$$BIC_{\Psi} = -2\mathcal{L}_n(\boldsymbol{\theta}) + (\tilde{p} + \tilde{q} - 1)\log n, \quad (28)$$

where $\tilde{p}$ is the dimensionality of $\boldsymbol{\beta}$ (*i.e.* the total number of non-zero regression coefficients in these model), and $\tilde{q}$ equal to $3K$ (FM-AR models) or $5K$ (for FM-GARCH models).

The Block-wise MM algorithm is iterated until some convergence criterion is met. In this article, we choose to use the absolute convergence criterion, where TOL > 0 is a small tolerance constant from [8]. Based on the discussion above, we summarise our algorithm in 1.

5. Simulated Data Analysis

In this section, we evaluate the performance of the proposed method and algorithm via simulations. We consider the Gaussian FM-AR models and Gaussian FM-GARCH models. Following [2] and [8], we used the correctly estimated zero coefficients (S1), correctly estimated non-zero coefficients (S2) and the mean estimate over all falsely identified non-zero predictors (M_{NZ}). The selection of thresholding parameters and components are solving by using Simulated Annealing (SA) algorithm. All simulations were evaluated with varying values of dimension p with 100 repetitions done for each.

5.1. Simulated Data Analysis of Gaussian FM-AR

The first simulations are based on the Gaussian FM-AR (2) model. Assuming that K is known, the model for the simulation was a $K = 2$ and $d = 2$ model of

$$\pi N\left(\alpha_1 + \mathbf{x}_t^{\mathrm{T}}\boldsymbol{\beta}_{11} + \mathbf{x}_{t-1}^{\mathrm{T}}\boldsymbol{\beta}_{12}, \sigma_1^2\right) + (1-\pi) N\left(\alpha_2 + \mathbf{x}_t^{\mathrm{T}}\boldsymbol{\beta}_{21} + \mathbf{x}_{t-1}^{\mathrm{T}}\boldsymbol{\beta}_{22}, \sigma_2^2\right), \quad (29)$$

Algorithm 1

Require:
Initialize $\boldsymbol{\lambda}^{(0)}$, $\boldsymbol{v}^{(0)}$, $\boldsymbol{a}^{(0)}$, $\boldsymbol{b}^{(0)}$, K, TOL;
Initialize $\boldsymbol{\pi}^{(0)}$, $\boldsymbol{\alpha}^{(0)}$, $\boldsymbol{\gamma_0}^{(0)}$, $\boldsymbol{\gamma}^{(0)}$, $\boldsymbol{\delta}^{(0)}$, $\boldsymbol{\beta}^{(0)}$;

Ensure:
Estimated value of the optimal solution $\boldsymbol{\pi}^{(*)}$, $\boldsymbol{\alpha}^{(*)}$, $\boldsymbol{\gamma_0}^{(*)}$, $\boldsymbol{\gamma}^{(*)}$, $\boldsymbol{\delta}^{(*)}$, $\boldsymbol{\beta}^{(*)}$ with appropriate $\boldsymbol{\lambda}^{(*)}$, $\boldsymbol{v}^{(*)}$, $\boldsymbol{a}^{(*)}$, $\boldsymbol{b}^{(*)}$, and given K;

```
while BIC_Ψ^(t+1) < BIC_Ψ^(t) do
  Update of λ^(t), v^(t), a^(t), b^(t).
  while | F_n^(r+1)(θ) − F_n^(r)(θ) | < TOL do
    Using Brent's algorithm in [8] with reverse communication to find unique root of (19).
    Update π^(r+1) by (18).
    Update α^(r+1) and γ0^(r+1) by soving (15).
    Update γ^(r+1) and δ^(r+1) by soving (16).
    Update β^(r+1) by soving (17).
    Update F_n^(r+1)(θ) by (9).
  end while
  Update BIC_Ψ^(t+1) by (28).
end while
```

where $n=300$, $p=10,20,100$, $\pi=0.3$, $\alpha_1=1$, $\alpha_2=5$, $\sigma_1=1$, and $\sigma_2=1$. Columns of $\boldsymbol{x}$ are drawn from a multivariate normal, with mean 0, variance 1, and two correlation structures: $\rho_{ij}=cor\left(x_i,x_j\right)=0.5^{|i-j|}$. The regression coefficients are

$$\boldsymbol{\beta}_{11}=(1,0,0,3,0,\cdots,0)^{\mathrm{T}}, \boldsymbol{\beta}_{12}=(-3,0,-1,0,2,\cdots,0)^{\mathrm{T}};$$

$$\boldsymbol{\beta}_{21}=(-1,2,0,0,3,\cdots,0)^{\mathrm{T}}, \boldsymbol{\beta}_{22}=(0,0,3,0,-2,\cdots,0)^{\mathrm{T}}.$$

Table 1 reports the results. We can see that when the dimension p = 100, the S2 in com1 of X_{t-1} from MIXSCAD-ML$_2$ is 100, however, the S2 in com1 of X_{t-1} from MIXLASSO-ML$_2$ (S2 = 70.7) and MIXMCP-ML$_2$ (S2 = 51.3) model are small, which indicates that MIXSCAD-ML$_2$ ensures that non-zero coefficients can be correctly identified and some non-zero coefficients in the MIXLASSO-ML$_2$ and MIXMCP-ML$_2$ model are not estimated. The mean estimate over all falsely identified non-zero predictors (M_{NZ}) of $\boldsymbol{\beta}$ from MIXSCAD-ML$_2$ are between 0.001 and 0.01.

5.2. Simulated Data Analysis of Gaussian FM-GARCH

The second simulations are based on the Gaussian FM-GARCH(2,1,1) model. Also assuming that K is known, the model for the simulation was a $K=2$, $d=2$, $M=1$ and $S=1$ model of

$$\pi N\left(\alpha_1 + \mathbf{x}_t^{\mathrm{T}}\boldsymbol{\beta}_{11} + \mathbf{x}_{t-1}^{\mathrm{T}}\boldsymbol{\beta}_{12}, \sigma_{1,t}^2\right) + (1-\pi) N\left(\alpha_2 + \mathbf{x}_t^{\mathrm{T}}\boldsymbol{\beta}_{21} + \mathbf{x}_{t-1}^{\mathrm{T}}\boldsymbol{\beta}_{22}, \sigma_{2,t}^2\right), \quad (30)$$

$$\sigma_{kt}^2 = \gamma_{0k} + \gamma_{k1}\epsilon_{k,t-1}^2 + \delta_{k1}\sigma_{k,t-1}^2, \quad (31)$$

for $k=1,2$, where $n=300$, $p=10,20$, $\pi=0.3$, $\alpha_1=2$ and $\alpha_2=5$, $\gamma_{01}=1$ and $\gamma_{02}=1$, $\gamma_{11}=0.5$ and $\gamma_{21}=0.2$, $\delta_{11}=0.4$ and $\delta_{21}=0.6$. $\epsilon_{kt}=\sigma_{kt}e_{kt}$, e_{kt} is an independent and identically distributed series with mean zero and variance unity. Columns of $\boldsymbol{x}$ are drawn from a multivariate normal, with mean 0, variance 1, and two correlation structures: $\rho_{ij}=cor\left(x_i,x_j\right)=0.5^{|i-j|}$. The regression coefficients are

Table 1. Summary of MIXLASSO-ML_2, MIXMCP-ML_2 and MIXSCAD-ML_2-penalized FM-AR (2) model with BIC method form the simulated scenario. Average correctly estimated zero coefficients (specificity; S_1), average correctly estimated non-zero coefficients (sensitivity; S_1), and the mean β estimate over all incorrectly estimated non-zero coefficients (M_{NZ}) are also reported.

Method	$K*d*p$	Com	X_t			X_{t-1}		
			$S_1(\%)$	$S_2(\%)$	M_{NZ}	$S_1(\%)$	$S_2(\%)$	M_{NZ}
MIXSCAD-ML_2	2*2*10	com1	86.0	99.5	0.097	90.0	99.7	−0.012
	2*2*20		91.2	99.5	0.067	91.6	99.7	−0.003
	2*2*100		81.7	100.0	0.016	82.6	100.0	0.009
		com2	94.3	99.3	0.020	95.5	100.0	−0.093
			94.2	99.3	0.013	96.1	100.0	−0.018
			90.7	100.0	-0.015	90.5	100.0	0.008
MIXMCP-ML_2	2*2*10	com1	80.1	100.0	0.040	87.6	100.0	0.005
	2*2*20		91.9	100.0	0.100	92.8	100.0	0.027
	2*2*100		98.1	81.0	0.304	98.1	51.3	0.205
		com2	93.0	100.0	0.041	96.5	100.0	−0.015
			96.8	100.0	0.055	98.4	100.0	0.084
			97.4	100.0	0.076	97.2	100.0	0.037
MIXLASSO-ML_2	2*2*10	com1	76.1	100.0	0.089	76.3	99.7	−0.019
	2*2*20		81.6	100.0	0.066	81.4	100.0	−0.011
	2*2*100		80.5	76.0	0.053	81.1	70.7	0.041
		com2	85.1	100.0	0.015	88.3	100.0	−0.001
			91.2	87.3	0.001	90.8	100.0	−0.015
			79.1	99.3	0.048	87.1	100.0	−0.039

$$\boldsymbol{\beta}_{11}=(1.5,0,2.5,0,0,\cdots,0)^{\mathrm{T}},\boldsymbol{\beta}_{12}=(-3.5,0,-1,0,2,\cdots,0)^{\mathrm{T}};$$

$$\boldsymbol{\beta}_{21}=(-1,2,0,0,3,\cdots,0)^{\mathrm{T}},\boldsymbol{\beta}_{22}=(0,0,3,0,-2,\cdots,0)^{\mathrm{T}}.$$

From **Table 2**, we can see that in all simulations, the value of S1 in com1 and com2 of X_t and X_{t-1} from MIXSCAD-ML_2 are the biggest, which indicates that MIXSCAD-ML_2 perform better than MIXLASSO-ML_2 and MIXMCP-ML_2 in correctly estimated zero coefficients. The mean estimate over all falsely identified non-zero predictors (M_{NZ}) of $\boldsymbol{\beta}$ from MIXSCAD-ML_2 is smaller than which from MIXLASSO-ML_2 and MIXMCP-ML_2.

6. Real Data Analysis

In this section, we evaluate the performance of the proposed method and algorithm via the analysis of the behavior of urban vehicular traffic of the city of São Paulo. This data set were collected notable occurrences of traffic in the metropolitan region of São Paulo in the period from 14 to 18 December 2009. This was acquired from the website http://archive.ics.uci.edu/ml/datasets.php. Registered from 7:00 to 20:00 every 30 minutes. It contains 135 observations and 18

Table 2. Summary of MIXLASSO-ML_2, MIXMCP-ML_2 and MIXSCAD-ML_2-penalized FM-GARCH(1, 1) model with BIC method form the simulated scenario. Average correctly estimated zero coefficients (specificity; S_1), average correctly estimated non-zero coefficients (sensitivity; S_1), and the mean β estimate over all incorrectly estimated non-zero coefficients (M_{NZ}) are also reported.

Method	$K*d*p$	Com	X_t			X_{t-1}		
			S_1 (%)	S_2 (%)	M_{NZ}	S_1 (%)	S_2 (%)	M_{NZ}
MIXSCAD-ML_2	2*2*10	com1	88.8	89.5	0.408	92.4	84.0	−0.048
	2*2*20		89.9	84.5	0.432	91.5	79.0	0.168
		com2	94.9	96.3	0.051	97.0	98.0	−0.139
			96.3	92.0	0.076	95.7	95.0	0.008
MIXMCP-ML_2	2*2*10	com1	80.8	94.0	0.417	87.4	81.3	0.115
	2*2*20		85.8	78.5	0.540	87.3	68.0	0.031
		com2	89.4	95.7	0.158	94.0	99.0	0.138
			93.4	91.0	0.269	95.6	95.5	0.118
MIXLASSO-ML_2	2*2*10	com1	73.9	84.5	0.426	79.9	76.0	−0.015
	2*2*20		81.3	66.5	0.579	83.5	56.7	−0.117
		com2	76.7	96.0	0.080	83.6	99.5	0.018
			88.2	75.0	0.111	93.4	90.5	−0.126

variables as well as one response variable. Covariate acronyms are hour (HO), immobilized bus (IB), broken truck (BT), vehicle excess (VE), accident victim (AV), running over (RO), fire vehicles (FV), occurrence involving freight (OIF), incident involving dangerous freight (IIDF), lack of electricity (LOE), fire (FI), point of flooding (POF), manifestations (MA), defect in the network of trolleybuses (DNT), tree on the road (TRR), semaphore off (SO), intermittent Semaphore (IS) and the response is slowness in traffic percent. Consider the effect of date on the behavior of traffic, we add a new variable that is day (DA). Figure 1 shows the heterogeneity of the data set, and the FM-AR or FM-GARCH model is applicable.

The levels of the covariates attributes from FMR, FM-AR (2) and FM-GARCH (2,1,1) with $K=2$ models are given in Table 3. From Table 4, we can see that the MIXSCAD-ML_2 penalized FM-GARCH (2,1,1) with $K=2$ model had the lowest BIC (622.9) across all analyses, the FM-AR (2) with $K=2$ model being ranked second ($\text{BIC}=677.3$), which is lower than the BIC (682.3) of FMR model. The predicted slowness in traffic percent from the FM-GARCH $K=2$ model had a MSE of 1.93 and a regression $\tilde{R}^2$ of 0.90. The predicted slowness in traffic percent from the FM-AR (2) $K=2$ model had a MSE of 2.09 and a regression $\tilde{R}^2$ of 0.89. The predicted slowness in traffic percent from the FMR $K=2$ model had a MSE of 2.41 and a regression $\tilde{R}^2$ of 0.87. These results suggest that the FM-GARCH (2,1,1) model had the smallest MSE and explained the largest proportion of variance for the slowness in traffic percent data. The results of the predicted response from these models are presented in Figure 2.

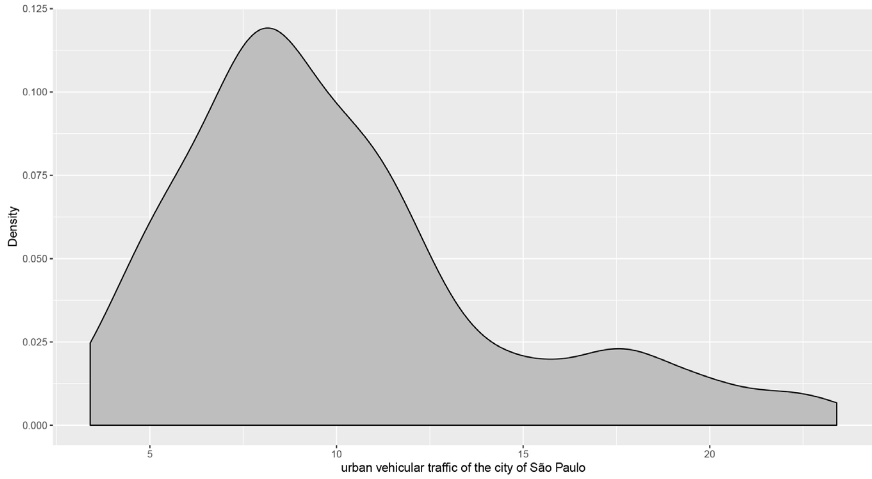


Figure 1. Density of slowness in traffic percent in the metropolitan region of São Paulo in the period from 14 to 18 December 2009.

Table 3. Summary of FMR, FM-AR and FM-GARCH model with BIC method and MIXLASSO-ML₂ penality.

Covariates	FMR		FM-AR				FM-GARCH			
	*com*1	*com*2	*com*1		*com*2		*com*1		*com*2	
			x_t	x_{t-1}	x_t	x_{t-1}	x_t	x_{t-1}	x_t	x_{t-1}
Intercept	7.32	−2.31	7.56	-	−1.89	-	1.39	-	6.24	-
π	0.37	0.63	0.34	-	0.66	-	0.47	-	0.53	-
DA	-	1.47	-	-	-	1.54	-	-	0.99	-
HO	0.13	0.52	0.11	0.36	-	0.13	0.29	0.39	-	−0.03
IB	-	-	-	-	-	-	-	-	-	-
BT	-	-	-	-	-	-	-	-	-	-
VE	-	-	-	-	-	-	-	-	-	-
AV	-	-	-	-	-	-	-	-	-	-
RO	-	-	-	-	-	-	-	-	-	-
FV	-	-	-	-	-	-	-	-	-	-
OIF	-	-	-	-	-	-	-	-	-	-
IIDF	-	-	-	-	-	-	-	-	-	-
LOE	-	1.75	-	-	-	1.88	-	-	-	1.80
FI	-	-	-	-	-	-	-	-	-	-
POF	-	0.61	-	1.25	-	-	-	1.41	-	-
MA	-	-	-	-	-	-	-	-	-	-
DNT	-	-	-	−0.91	-	-	-	−0.71	-	-
TRR	-	-	-	-	-	-	-	-	-	-
SO	-	-	-	-	-	-	-	-	-	-
IS	-	-	-	-	-	-	-	-	-	-

Table 4. Summary of the values of BIC, MSE, and adjusted regression (predicted response on observed response) $\tilde{R}^2$ from FMR, FM-AR (2) and FM-GARCH (2,1,1) models.

model	K	BIC	MSE	$\tilde{R}^2$
FM-GARCH (2,1,1)	2	622.90	1.93	0.90
FM-AR (2)	2	677.32	2.09	0.8
FMR	2	682.36	2.41	0.87

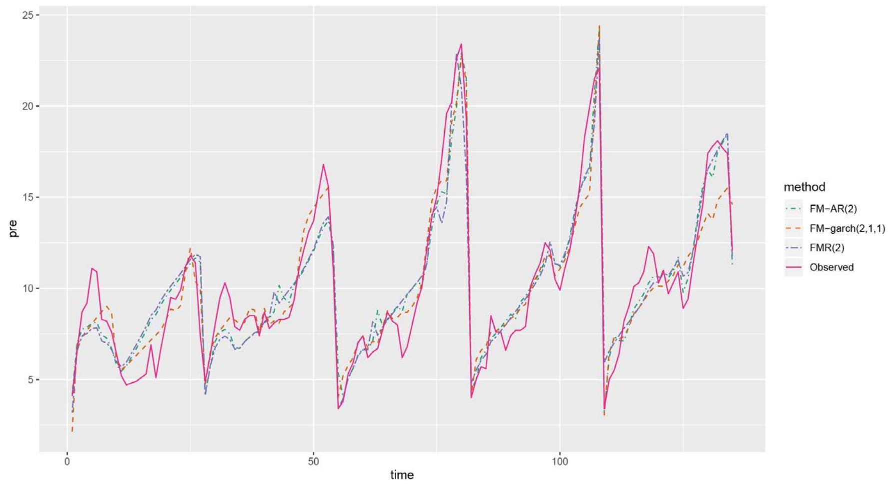


Figure 2. Summary of predicted and observed slowness in traffic percent in the metropolitan region of São Paulo in the period from 14 to 18 December 2009.

7. Discussion

In this article, we disccused that the modeling of response variable which is time series and with a finite mixture distribution depends on covariates, and the variable selection problem of them. We propose the FM-AR models and FM-GARCH models for modeling data that arise from a heterogeneous population which is time series, and propose a new regularization method (MIXLASSO-ML$_2$, MIXMCP-ML$_2$, MIXSCAD-ML$_2$) for the variable selection in these model, which composed of the mixture of the l_1 penalty and l_2 penalty proportional to mixing proportions. In addition, we estimate the maximum log quasi-likelihood estimate for the new penalized FM-AR and FM-GARCH model, and derive a general expression for the block-wise minimized maximization (MM) algorithm with better features. The simulation results of Gaussian FM-AR and Gaussian FM-GARCH models and an actual data set illustrate the capability of the methodology and algorithm, and MIXSCAD-ML$_2$ is always superior to other penalty methods.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] McLachlan, G.J. and Peel, D. (2000) Finite Mixture Models. Wiley, New York. https://doi.org/10.1002/0471721182

[2] Abbas, K. and Shili, L. (2013) Regularization in Finite Mixture of Regression Models with Diverging Number of Parameters. *Biometrics*, **69**, 201-235. https://doi.org/10.1111/biom.12020

[3] Abbas, K. and Chen, J.H. (2007) Variable Selection in Finite Mixture of Regression Models. *Journal of the American Statistical Association*, **102**, 1025-1038. https://doi.org/10.1198/016214507000000590

[4] Zhang, Z., Li, W.K. and Yuen, K.C. (2005) On a Mixture GARCH Time-Series Model. *Journal of Time Series Analysis*, **27**, 577-597. https://doi.org/10.1111/j.1467-9892.2006.00467.x

[5] Haas, M., Krause, J., Paolella, M.S. and Steude, S.C. (2013) Time-Varying Mixture GARCH Models and Asymmetric Volatility. *North American Journal of Economics and Finance*, **26**, 602-623. https://doi.org/10.1016/j.najef.2013.02.024

[6] Abbas, K. (2011) An Overview of the New Feature Selection Methods in Finite Mixture of Regression Models. *Journal of the Iranian Statistical Society*, **10**, 201-235.

[7] Fan, J.Q. and Li, R.Z. (2001) Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, **96**, 1348-1360. https://doi.org/10.1198/016214501753382273

[8] Lloyd-Jones, L.R., Nguyen, H.D. and McLachlan, G.J. (2018) A Globally Convergent Algorithm for Lasso-Penalized Mixture of Linear Regression Models. *Computational Statistics and Data Analysis*, **119**, 19-38. https://doi.org/10.1016/j.csda.2017.09.003

[9] Hwang, S.Y., *et al.* (2014) Quasilikelihood and Quasi-Maximum Likelihood for GARCH-Type Processes: Estimating Function Approach. *Journal of the Korean Statistical Society*, **1**, 5. https://doi.org/10.1016/j.jkss.2014.01.005

[10] Razaviyayn, M., Hong, M. and Luo, Z.-Q. (2013) A Unified Convergence Analysis of Block Successive Minimization Methods for Nonsmooth Optimization. *SIAM Journal on Optimization*, **23**, 1126-1153. https://doi.org/10.1137/120891009

[11] Lange, K. (2013) Optimization. Springer, New York. https://doi.org/10.1007/978-1-4614-5838-8

Advances in Pure Mathematics, 2020, 10, 114-124
https://www.scirp.org/journal/apm
ISSN Online: 2160-0384
ISSN Print: 2160-0368

On Conditional Probabilities of Factoring Quadratics

Thomas Beatty, Gabriela von Linden

Department of Mathematics, Florida Gulf Coast University, Fort Myers, FL, USA
Email: tbeatty@fgcu.edu, gaby.vonlinden@gmail.com

How to cite this paper: Beatty, T. and von Linden, G. (2020) On Conditional Probabilities of Factoring Quadratics. *Advances in Pure Mathematics*, **10**, 114-124.
https://doi.org/10.4236/apm.2020.103008

Received: January 30, 2020
Accepted: March 16, 2020
Published: March 19, 2020

Abstract

Factoring quadratics over $\mathbb{Z}$ is a staple of introductory algebra and textbooks tend to create the impression that doable factorizations are fairly common. To the contrary, if coefficients of a general quadratic are selected randomly without restriction, the probability that a factorization exists is zero. We achieve a specific quantification of the probability of factoring quadratics by taking a new approach that considers the absolute size of coefficients to be a parameter n. This restriction allows us to make relative likelihood estimates based on finite sample spaces. Our probability estimates are then conditioned on the size parameter n and the behavior of the conditional estimates may be studied as the parameter is varied. Specifically, we enumerate how many formal factored expressions could possibly correspond to a quadratic for a given size parameter. The conditional probability of factorization as a function of n is just the ratio of this enumeration to the total number of possible quadratics consistent with n. This approach is patterned after the well-known case where factorizations are carried out over a finite field. We review the finite field method as background for our method of dealing with $\mathbb{Z}[x]$. The monic case is developed independently of the general case because it is simpler and the resulting probability estimating formula is more accurate. We conclude with a comparison of our theoretical probability estimates with exact data generated by a computer search for factorable quadratics corresponding to various parameter values.

Keywords

Factorization, Polynomial, Quadratic, Integers, Rational Numbers, Monic, Modular Arithmetic, Conditional Probability

1. Introduction

This paper presents the preliminary results of a broader program to estimate the probabilities of factoring more general polynomials over $\mathbb{Z}$. We anticipate that subsequent research will develop along the lines suggested by the quadratic case investigated here, specifically by using the method of parameterizing the maximum absolute value of coefficients and correlating the conditional probabilities of factorization with the size of the parameter.

The probability that a given general quadratic $\alpha x^2+\beta x+\gamma\in\mathbb{Z}[x]$ can be factored depends heavily on the commensurability of the coefficients [1] [2] [3]. Loosely speaking, if the coefficients are all about the same size in absolute value, and that size is small, the existence of a factorization is relatively more likely than otherwise. Our intention is to quantify this phenomenon. Dealing with the infinite number of choices available for coefficients is a problem. We sidestep this obstacle by adapting the method used to determine the probability of factorization of quadratics over finite fields. Briefly, we establish a cutoff, or size parameter *n*, for the absolute value of any coefficients appearing in any of the quadratics we wish to study. This makes the number of quadratics under consideration finite as well as the number of formal factored expressions that could possibly yield such a quadratic. Then the classical probability is just the ratio of the number of admissible factored expressions to the total number of quadratics which conform to the cutoff. This probability $P(n)$ is, of course, a conditional probability given that the coefficients do not exceed *n* in absolute value. So the infinite character of the problem is initially made finitary where calculations can be done and then can be recovered by allowing *n* to approach infinity.

We consider three cases, the first of which, factoring quadratics over finite fields, is well-known [4]. For factoring quadratics over $\mathbb{Z}$, we split the discussion into two parts: 1) the monic case, and 2) the general case. For simplicity we consider quadratics in $\mathbb{Z}[x]$ that have non-negative roots.

Figure 1 shows factorization probabilities calculated by the computer search. Currently we have no formula that estimates the case $ax^2+bx+c\in\mathbb{Z}[x]$ with $\alpha\neq 0$ other than curve fitting. The trend in the graph in **Figure 1** is borne out by the following proposition.

Proposition 0

If $a,b,c\in\mathbb{Z}$ are selected randomly without restriction, then the probability of factoring ax^2+bx+c over $\mathbb{Z}$ is zero.

Proof. Suppose we are given ax^2+bx+c with $a,b,c\in\mathbb{Z}$ selected at random. This quadratic factors over $\mathbb{Q}$, hence $\mathbb{Z}$ by Gauss' Lemma, if the discriminant $\Delta=b^2-4ac$ is a perfect square. We ask what is the probability that Δ is a perfect square if *a* and *b* have been selected and *c* is provisionally allowed to range over $[-n,n]$ for some $n\in\mathbb{N}$. Then there are $2n+1$ possible values of Δ spread over an interval of length $|8an|$. The largest possible number of perfect squares in such an interval would occur when none of the interval intersected the open left half line and where the arithmetic density of squares was the greatest. This would occur if the interval were exactly $[0,|8an|]$. The number of squares in this interval does not exceed $\sqrt{|8an|}$. The classical probability that

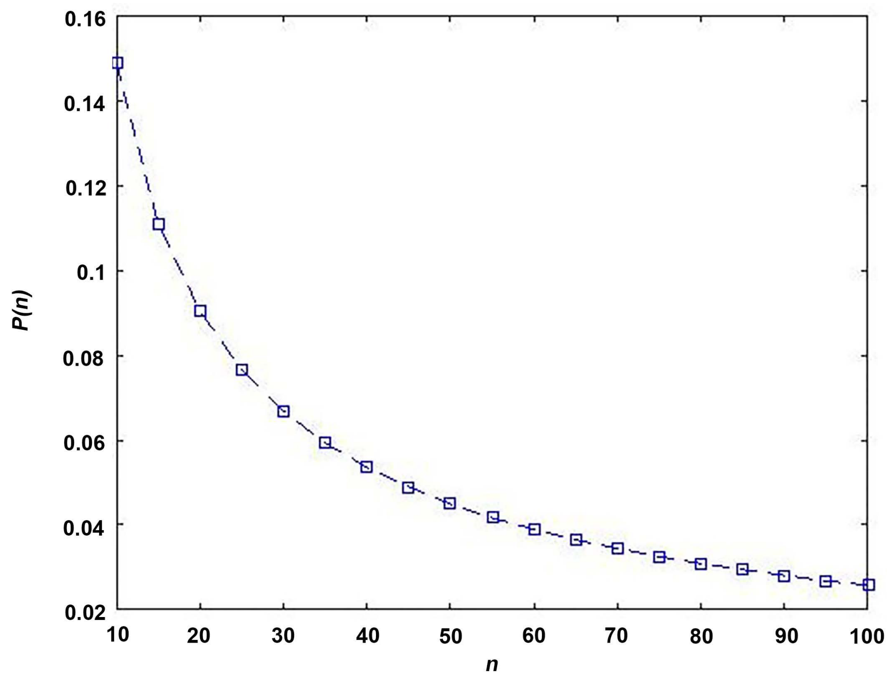


Figure 1. $P(n)$ is the probability of factoring ax^2+bx+c with $|a|,|b|,|c| \le n$.

one of these squares coincides with a value of Δ is therefore no more than $\frac{\sqrt{|8an|}}{2n+1} < \sqrt{2}|a|\frac{1}{\sqrt{n}}$. As the provisional restriction that $c \in [-n,n]$ is relaxed by allowing $n \to \infty$, we see $\sqrt{2}|a|\frac{1}{\sqrt{n}} \to 0$. It follows that the probability that Δ is a perfect square, and therefore ax^2+bx+c is factorable over $\mathbb{Z}$, is zero in the limit, which corresponds to no restrictions at all on *c*. Since this is true for any triple (a,b,c), the proposition is established.

We wish to have a more granular understanding of the way in which factorability depends on commensurability of coefficients. Our approach to this question is motivated by solving the factorization probability problem in the context of finite fields, which we review below.

2. Factoring Over $GF(p^n)$

Suppose we are given a random monic quadratic over the finite field $GF(p)$. What is the likelihood that it factors [5]?

2.1. Proposition 1

If $f(x)=x^2+\alpha x+\beta$ and $\alpha,\beta \in GF(p)$, then the probability P_p of factoring $f(x)$ over $GF(p)$ is $\frac{1}{2}+\frac{1}{2p}$.

Proof. There are p^2 possible pairs of coefficients, hence p^2 distinct quadratics. On the other hand, if *f* factors as $(x-r_1)(x-r_2)$, then there are *p* factorizations where $r_1 = r_2$ and $\binom{p}{2}$ distinct factorizations where $r_1 \neq r_2$, allowing for interchanging the factors. P_p is the ratio of possible factorizations to possible quadratics, so $P_p = \dfrac{p + \dfrac{p(p-1)}{2}}{p^2} = \dfrac{1}{2} + \dfrac{1}{2p}$.

Now let us generalize to an arbitrary quadratic.

2.2. Corollary 1-1

If $f(x) = \lambda x^2 + \alpha x + \beta$ and $\lambda(\neq 0), \alpha, \beta \in GF(p)$, then the probability P_p of factoring $f(x)$ over $GF(p)$ is $\dfrac{1}{2} + \dfrac{1}{2p}$.

Proof. Evidently there are $(p-1)p^2$ possible triples of coefficients, but we can mimic the above proof by rewriting
$f(x) = \lambda(x^2 + \lambda^{-1}\alpha x + \lambda^{-1}\beta) = \lambda(x^2 + \alpha' x + \beta')$. Now the possible factorizations would look like $\lambda(x - r_1)(x - r_2)$. Once again, the probability of factoring a random quadratic, not necessarily monic, would be

$$P_p = \frac{(p-1)\left(\dfrac{p^2 + p}{2}\right)}{(p-1)p^2} = \frac{1}{2} + \frac{1}{2p}.$$

2.3. Corollary 1-2

If $f(x) = \lambda x^2 + \alpha x + \beta$ and $\lambda(\neq 0), \alpha, \beta \in GF(p^n)$, then the probability P_{p^n} of factoring $f(x)$ over $GF(p^n)$ is $\dfrac{1}{2} + \dfrac{1}{2p^n}$.

Proof. Following the proof for the preceding, we have $(p^n - 1)p^{2n}$ possible triples of coefficients, and $(p^n - 1)\left[p^n + \dfrac{p^n(p^n - 1)}{2}\right] = (p^n - 1)\left(\dfrac{p^{2n} + p^n}{2}\right)$ possible factorizations. Hence $P_{p^n} = \dfrac{(p^n - 1)\left(\dfrac{p^{2n} + p^n}{2}\right)}{(p^n - 1)p^{2n}} = \dfrac{1}{2} + \dfrac{1}{2p^n}$.

2.4. Corollary 1-3

The limit as $p \to \infty$ of the probability P_{p^n} of factoring a quadratic over $GF(p^n)$ is $\dfrac{1}{2}$.

Proof. This follows immediately from the fact that the limit as $p \to \infty$ of the expression in Corollary 1-2 is independent of *n*.

The situation we see embodied in Proposition 1 and its corollaries is somewhat unexpected (at least the first time it is considered) and in any case very different from factoring over $\mathbb{Z}$. It is a mildly entertaining exercise in experimental mathematics to choose a large prime p and ask a computer algebra system to factor several random quadratics with large coefficients modulo p. Superficially, since p is large, it seems that the chances for a factorization would be about the same as if the factorization were to be done over $\mathbb{Z}$, namely poor. But in the long run, about half of the test examples result in factorizations. Although it would defeat the whole purpose of our discussion of the enumeration method for factorization probabilities in the finite field case, a short proof of Proposition 1 can be gotten directly from number theory. The quadratic equation $x^2+\alpha x+\beta=0$ can be simplified by completing the square, leaving a constant on the right hand side. Among the $p-1$ non-zero least residues modulo p, exactly $\frac{p-1}{2}$ are quadratic residues. The constant needs to be a quadratic residue so that roots can be found to construct a factorization. Zero is always a quadratic residue, so there are $\frac{p-1}{2}+1=\frac{p+1}{2}$ quadratic residues in all. On the other hand, there are p least residues modulo p, hence the probability that a randomly chosen least residue is in fact a quadratic residue is $\frac{p+1}{2p}=\frac{1}{2}+\frac{1}{2p}$ as above.

3. Factoring Over $\mathbb{Z}$-Monic Case

We would like to adapt the same argument for factoring over $\mathbb{Z}$ as we used for finite fields, namely counting up the number of possible distinct factorizations and dividing by the number of distinct quadratic expressions to get a probability of being able to factor. Since $\mathbb{Z}$ is infinite this plan is immediately hobbled [6]. Consider the polynomial x^2+ax+1, where α is random. There are only two hopes for factorization: $\alpha=\pm 2$. Yet there are infinitely many choices for α, so the probability of factorization is evidently zero. To salvage any insight from this state of affairs, we have to content ourselves with a conditional probability based on limiting how "random" a random quadratic can be. A reasonable choice is to insist that its coefficients be commensurable with its possible zeroes. Clearly $x^2+37x+1$ does not have this property, which is informal at the moment, but which will soon be made precise. On the other hand, both x^2+x+1 and x^2+2x+1 seem to have it. In one case there is a factorization, in the other not. This conditional probability will be based on a relative likelihood calculation where the commensurability condition is quantified by a parameter. As the parameter increases, the commensurability decreases, and the probability of factorization will tend to zero. The point of our approach is to model the detailed behavior of this process. To introduce the specifics in a simple context, we first consider monic quadratics with the restriction that they have non-negative roots.

Let us define a "window of feasibility" $W(n)$ in the plane as the rectangle of grid points $[0,n+1]\times[0,n]\subset\mathbb{Z}^2$. Suppose we are given a random $p(x)\in\mathbb{Z}[x]$ of the form $x^2-\alpha x+\beta$ with $\alpha,\beta\geq 0$ and $\beta\leq n$. If $p(x)$ factors as

$(x-r_1)(x-r_2)$ both roots are nonnegative and $\alpha = r_1 + r_2$ and $\beta = r_1 r_2$. If $\beta > 0$ these conditions imply that $\alpha \le n+1$, and in the event $\beta = 0$ we impose this condition arbitrarily to ensure that all $p(x)$ satisfying these conditions can be mapped in the obvious way to $(\alpha,\beta) \in W(n)$. At the moment we have every point in $W(n)$ as the image of some $p(x)$ with $\alpha,\beta \ge 0$. Clearly this is a bijective mapping. We call $W(n)$ a window of feasibility since any $p(x)$ with $0 < \beta \le n$ and $\alpha > n+1$ is necessarily impossible to factor. The motivation for constructing the window is to exclude those cases which overwhelm ordinary probability calculations. A quadratic inside the window may or may not be factorable, but if not, the reason will not be due to incommensurability of coefficients. Figure 2 shows this situation.

Note that a grid point of the form $(\alpha,0)$ corresponds to the quadratic $x^2 - \alpha x = x(x-\alpha)$, so factorization is always possible. A grid point of the form $(0,\beta)$, provided $\beta > 0$ corresponds to the quadratic $x^2 + \beta$, which never factors.

Now denote by $F(n)$ the grid points in $W(n)$ corresponding to factorable polynomials. A factorable quadratic mapped to (α,β) must have $\alpha = r_1 + r_2$ and $\beta = r_1 r_2$ for some r_1, r_2. Let us then plot pairs of roots on the $W(n)$ grid and define $\phi(r_1,r_2) = (\alpha,\beta)$ as the mapping that associates a root pair to the quadratic which factors with those roots. We will see shortly that there are substantial limitations on which points in $W(n)$ can be root pairs. In other words,

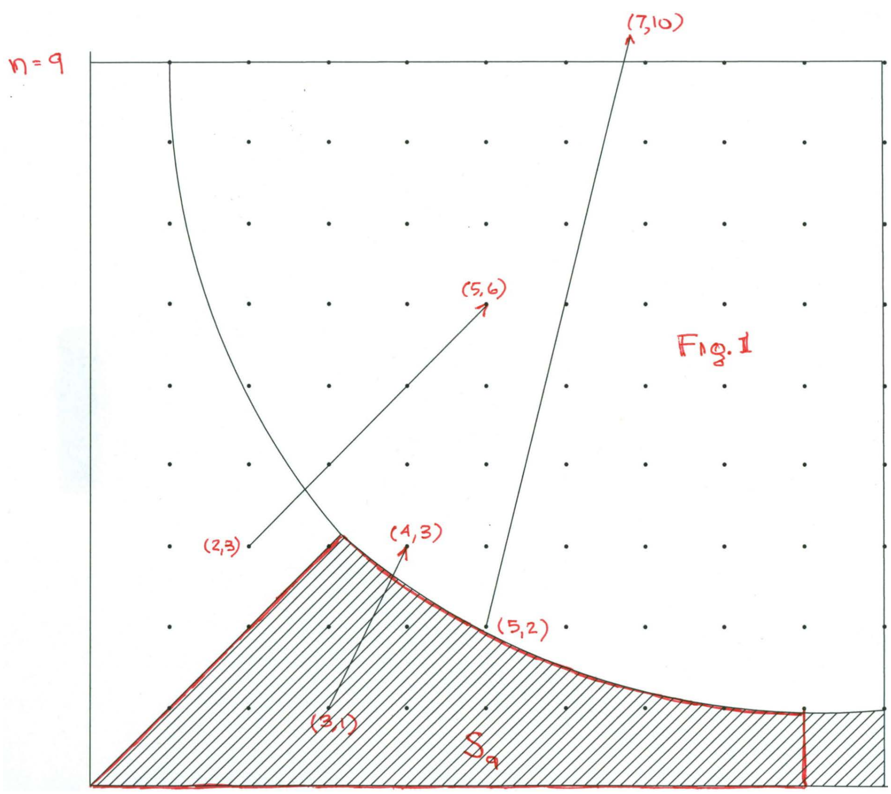


Figure 2. Window of Feasibility for $n=9$. Grid points in shaded area are possible root pairs.

the domain of ϕ will be relatively small, so surjectivity will clearly be out of the question. We would like ϕ to be injective, and since $\phi(r_1,r_2)=\phi(r_2,r_1)$ we impose the condition $r_2 \le r_1$ without loss of generality. Graphically, this amounts to eliminating all root pairs strictly above the line $r_2 = r_1$. Now $W(n)\setminus\{(r_1,r_2): r_2 \le r_1\}$ is still not $F(n)$ since there is another important condition on the root pairs, namely $r_1 r_2 \le n$. The only admissible root pair grid points whenever $r_2 > 0$ are also below the hyperbola $r_2 = \frac{n}{r_1}$. This is also true trivially if $r_2 = 0$ so the hyperbola is an upper bound for all root pairs corresponding to factorable quadratics. Now we estimate the number of points in $F(n)$ by calculating the area inside the region of the plane defined by the line $r_2 = 0$, the line $r_2 = r_1$, and the hyperbola $r_2 = \frac{n}{r_1}$. The area of this region is $A = \int_0^{\sqrt{n}} (r_1)\mathrm{d}r_1 + \int_{\sqrt{n}}^{n+1} \left(\frac{n}{r_1}\right)\mathrm{d}r_1 = \frac{n\ln n + 3n}{2}$. We estimate the number of grid points in $F(n)$ by assuming one grid point per unit of area. This estimate is asymptotically exact. Now $W(n)$ has $(n+1)(n+2)$ total grid points. Finally, our conditional probability estimate is $\frac{n\ln n + 3n}{2(n+1)(n+2)}$. We have proved:

3.1. Proposition 2

Given a random quadratic in $\mathbb{Z}[x]$ of the form $x^2 - \alpha x + \beta$, where $\alpha,\beta \ge 0$, the approximate probability $P(n)$ of factorization, subject to the conditions $\alpha \le n+1$ and $\beta \le n$ for fixed $n \in \mathbb{N}$, is given by $P(n) = \frac{n\ln n + 3n}{2(n+1)(n+2)}$.

3.2. Corollary 2-1

With the notation of Proposition 2, $P(n)$ is asymptotically $\frac{\ln n}{2n}$.

Proof. Divide numerator and denominator of $\frac{n\ln n + 3n}{2(n+1)(n+2)}$ by n to get $\frac{\ln n + 3}{2\left(n + 3 + \frac{2}{n}\right)}$. Note that for large n we have $\ln n \gg 3$ and $n \gg 3 + \frac{2}{n}$. Hence $n \gg 1$ implies $P(n) \approx \frac{\ln n}{2n}$.

Unsurprisingly, letting $n \to \infty$ corresponds to α and β being chosen completely arbitrarily and we confirm that $\lim_{n\to\infty} P(n) = 0$ by L'Hôpital's rule. The rate at which the likelihood of factorization declines is $P'(n) \approx -\frac{\ln n}{2n^2}$ for large n, as would be expected.

For perspective, if $n = 440$, the corresponding likelihood of factorability $P(440) \approx 1\%$.

4. Factoring over $\mathbb{Z}$-General Case

Consider the lattice cube $L(n)=[0,n]^3 \subset \mathbb{Z}^3$ as the three-dimensional analog of the preceding window of feasibility. The general quadratic [7] $p(x)=\alpha x^2-\beta x+\gamma \in \mathbb{Z}[x]$ with $\alpha>0,\beta,\gamma \geq 0$ can be associated injectively with the point $(\alpha,\beta,\gamma)\in L(n)$ provided $\max\{\alpha,\beta,\gamma\}\leq n$. Since $\alpha \neq 0$, there are $(n-1)n^2$ grid points in $L(n)$ which represent distinct quadratics of this type. We would like to estimate the number of these which are factorable so that we can calculate the conditional probability $P(n)$ of factorability in the manner of the finite field and monic cases above. Suppose $p(x)$ does factor in $\mathbb{Z}$ as $(ax-b)(cx-d)$, with $a,c>0$ and $b,d\geq 0$. Note that the greatest common divisor of α,β and γ does not appear as a separate factor but is considered to be bundled into the term $(ax-b)$. Then $\alpha = ac$, $\beta = bc+ad$, and $\gamma = bd$. Since the maximum α is *n*, it follows that admissible pairs (a,c) would satisfy $c \leq \frac{n}{a}$. As in the monic case, we embed $\mathbb{Z}^2$ in $\mathbb{R}^2$, calculate the appropriate area under the given hyperbola, and then assume one grid point (admissible pair) per unit of area with the understanding that this is asymptotically correct. The area in the first quadrant of the *ac* plane is $\int_1^n \frac{n(\delta a)}{a}-n=n\ln n-n=n(\ln n-1)$. Since the maximum γ is *n*, but either *b* or *d* (or both) could be zero, the appropriate area in the first quadrant of the *bd* plane would be $\int_1^n \frac{n(\delta b)}{b}+n=n\ln n+n=n(\ln n+1)$. Evidently there are $[n(\ln n-1)]\cdot[n(\ln n+1)]=n^2\left((\ln n)^2-1\right)=F(n)$ expressions of the form $(ax-b)(cx-d)$ which could factor the quadratic $\alpha x^2-\beta x+\gamma$ subject to the constraints imposed on those coefficients. Certainly not every point in $L(n)$ corresponds to a factorable quadratic, but we can be assured that any points which are associated with a factorable quadratic have a factorization counted by $F(n)$. Since the factor pairs commute, we divide $F(n)$ by two and it follows that the conditional probability of factorization is estimated by

$$P(n)=\frac{F(n)}{2n^3}=\frac{n^2\left((\ln n)^2-1\right)}{2n^3}=\frac{(\ln n)^2-1}{2n}.$$

We have proved:

4.1. Proposition 3

Given a random quadratic in $\mathbb{Z}[x]$ of the form $\alpha x^2-\beta x+\gamma$, where $\alpha>0,\beta,\gamma\geq 0$ and $\max\{\alpha,\beta,\gamma\}\leq n$, an estimate for the probability $P(n)$ of factorization is given by $P(n)=\frac{(\ln n)^2-1}{2n}$.

4.2. Corollary 3-1

With the notation of Proposition 3, $P(n)$ is asymptotically $\frac{(\ln n)^2}{n}$.

Table 1. Actual vs. Calculated-Monic Case.

n	Calc	Actual
10	0.200	0.197
20	0.128	0.121
30	0.097	0.089
40	0.078	0.070
50	0.065	0.058
1000	0.005	*
∞	0	*

n = coefficient bound; Calc = calculated $P(n)$; Actual = actual $P(n)$ by computer check.

Table 2. Actual vs. Calculated-General Case.

n	Calc	Actual
10	0.215	0.149
20	0.199	0.190
30	0.176	0.067
40	0.158	0.054
50	0.143	0.045
1000	0.023	*
∞	0	*

n = coefficient bound; Calc = calculated $P(n)$; Actual = actual $P(n)$ by computer check.

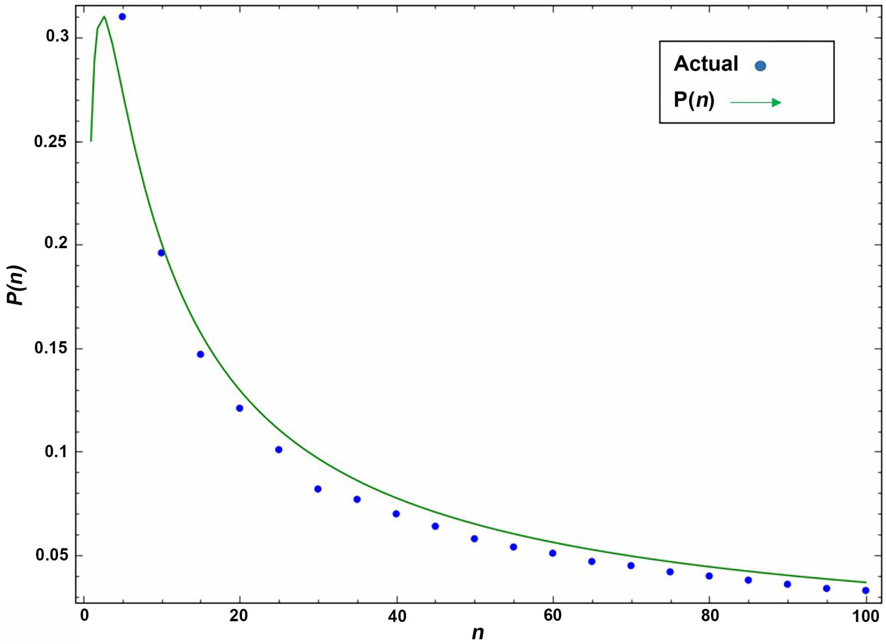


Figure 3. Probability of factoring $P(n)$ vs. absolute coefficient bound n actual vs. calculated.

Proof. For $n \gg e$, $\frac{(\ln n)^2 - 1}{2n} \approx \frac{(\ln n)^2}{2n}$.

As expected, $\lim_{n\to\infty} P(n) = 0$ again by L'Hôpital's rule.

5. Summary & Conclusions

We have described two methods for estimating the conditional probability that a random quadratic in $\mathbb{Z}[x]$ with non-negative bounded coefficients can be factored as a function of the bounding parameter. The simpler case is based on mapping monic quadratics injectively to a two-dimensional lattice in $\mathbb{Z}^2$ and enumerating the formal expressions that could possibly represent factorizations of them. The ratio of the number of admissible formal factorizations to the total number of points in the lattice defines the conditional probability of factorization for the given coefficient bound. The more complicated case involves mapping general quadratics to a three-dimensional lattice in $\mathbb{Z}^3$ and reprising the calculation for the two-dimensional case. Both methods have their provenance in the problem of calculating the likelihood that a quadratic over a finite field may be factored. In the case of finite fields, only a finite number of polynomials are possible and only a finite number of factorizations can be written, making the calculation a simple ratio. This fails, of course, for $\mathbb{Z}$, but the point of our method is to resurrect the utility of finiteness by imposing a size limitation on coefficients [8].

Table 1 presents a comparison of values from the monic formula for conditional probability given by Proposition 2 with a computer generated census of factorable monic quadratics. There is reasonably close agreement, even for small *n*. The computer algorithm works by simply checking to see if the quadratic formula yields a rational number. Recall if a polynomial in $\mathbb{Z}[x]$ factors over $\mathbb{Q}$, then it factors over $\mathbb{Z}$.

Table 2 recaps a similar comparison for the general quadratics in Proposition 3. For the sake of simplicity we have ignored double counting certain factored expressions arising from symmetries (for example if $a = c$). This overstates the calculated probability of factorization, especially for small *n*, so we may regard $P(n)$ in this case as an upper bound for the true probability. In any case, our formula establishes $P(\infty) = 0$.

Below in Figure 3 a graph of the calculated $P(n)$ versus *n* is shown as the continuous curve. The separate data points correspond to Table 1. The expected feature that $\lim_{n\to\infty} P(n) = 0$ is also evident.

To close on a philosophical note, although factorization of random quadratics over $\mathbb{Z}$ has been shown to be a progressively futile exercise, practicing pattern recognition with doable examples for small *n* is a worthwhile exercise that no doubt pays dividends elsewhere in mathematics.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Burton, D. (2005) Elementary Number Theory. McGraw-Hill Higher Education, Inc., New York, NY.

[2] Dummit, D.S. and Foote, R.M. (2014) Abstract Algebra. 3rd Edition, John Wiley and Sons, Inc., Hoboken, NJ.

[3] Gallian, J. (2010) Contemporary Abstract Algebra. 7th Edition, Cengage Learning.

[4] Morandi, P. (1996) Field and Galois Theory. Springer, Berlin. https://doi.org/10.1007/978-1-4612-4040-2

[5] Lenstra, A., Lenstra, H. and Lovász, L. (1982) Factoring Polynomials with Rational Coefficients. *Mathematische Annalen*, **261**, 515-534. https://doi.org/10.1007/BF01457454

[6] Miller, V.S. (1992) Factoring Polynomials via Relation-Finding. In: Dolev, D., Galil, Z. and Rodeh, M., Eds., *Theory of Computing and Systems ISTCS* 1992, Lecture Notes in Computer Science, Vol. 601, Springer, Berlin, Heidelberg.

[7] Hart, W., Hoeij, M. and Novocin, A. (2011) Practical Polynomial Factoring in Polynomial Time. *Proceedings of the International Symposium on Symbolic and Algebraic Computation* (*ISSAC*), San Jose, CA, 163-170. https://doi.org/10.1145/1993886.1993914

[8] Ostrowski, A.M. (1975) On Multiplication and Factorization of Polynomials I, *Aequationes Mathematicae*, **13**, 201-228. https://doi.org/10.1007/BF01836524

Scientific Research Publishing

Advances in Pure Mathematics, 2020, 10, 125-154
https://www.scirp.org/journal/apm
ISSN Online: 2160-0384
ISSN Print: 2160-0368

From Pythagoras Theorem to Fermat's Last Theorem and the Relationship between the Equation of Degree *n* with One Unknown

Yufeng Xia

Huaqiao University, Fujian, China
Email: qhdxxyf1991@163.com

How to cite this paper: Xia, Y.F. (2020) From Pythagoras Theorem to Fermat's Last Theorem and the Relationship between the Equation of Degree *n* with One Unknown. *Advances in Pure Mathematics*, **10**, 125-154.
https://doi.org/10.4236/apm.2020.103009

Received: August 27, 2019
Accepted: March 22, 2020
Published: March 25, 2020

Abstract

The most interesting and famous problem that puzzled the mathematicians all around the world is much likely to be the Fermat's Last Theorem. However, since the Theorem was proposed, people can't find a way to solve the problem until Andrew Wiles proved the Fermat's Last Theorem through a very difficult method called Modular elliptic curves in 1995. In this paper, I firstly constructed a geometric method to prove Fermat's Last Theorem, and in this way we can easily get the conclusion below: If a and b are integer and $a = b$, $n \in \mathbb{Q}$ and $n > 1$, the value of c satisfies the function $a^n + b^n = c^n$ that can never be integer; if a, b and c are integer and $a \neq b$, n is integer and $n > 2$, the function $a^n + b^n = c^n$ cannot be established.

Keywords

Pythagoras Theorem, Fermat's Last Theorem, Geometric Method, Equation of Degree n with One Unknown

1. Introduction

The Fermat's Last Theorem was proposed by French famous mathematician Pierre de Fermat in 1637, it was called the last theorem because it was the theorem of Fermat that can be proved at last, which means to prove the theorem is very difficult. The Fermat's Last Theorem states: there is no positive integer a, b and c to satisfy the function $a^n + b^n = c^n$ (1) when n is integer and $n > 2$ [1].

Many mathematicians paid attention to this theorem, and they found it not as easy as it looks like. In 1753, the famous Swiss mathematician Euler said in a letter to Goldbach that he proved the Fermat conjecture at $n = 3$, and his proof was published in the book Algebra Guide in 1770 [2]. Fermat himself proved the Fermat conjecture at $n = 4$ [3]. In 1825, the German mathematician Dirichlet and the French mathematician Legendre independently proved that Fermat's theorem was established at $n = 5$, using the extension of the method used by Eulerasin [2]. In 1844, Kummer proposed the concept of "ideal number", he proved that for all prime indices n less than 100, Fermat's theorem was established, and this study came to a stage [4]. But the mathematicians still struggled with Fermat's theorem in the first two hundred years of the conjecture with little progress. What's more, many theorem were proposed in order to prove the Fermat's Last Theorem, such as Modell conjecture, Taniyama-Shimura theorem. After proving the Taniyama-Shimura theorem, Andrew Wiles finally got a way to prove the Fermat's Last Theorem in 1995 [5].

At first, people wanted to prove that the Fermat's Last Theorem was established at different indices n, but the indices n is infinite, this method is meant to be failed. Then, people tried to propose another theorem to indirectly prove the Fermat's Last Theorem, but the relationship between two theorems is not very clear, thus the proof is hard to be verified.

To prove the Fermat's Last Theorem, I got inspiration from the Pythagoras Theorem. As we all know the Pythagoras Theorem: the sum of the squares of the two right-angled sides of a right-angled triangle is equal to the square of the hypotenuse, let the length of two right-angled sides be a and b, and the length of hypotenuse is c, then $a^2 + b^2 = c^2$ (2) [6]. What's more, if a, b and c satisfy the function (2), the angle (θ) between a and b must be $\frac{\pi}{2}$. If a, b and c satisfy the function $a^n + b^n = c^n$ (1), what is the relationship between n and θ? This paper discusses the relationship between n and θ, and in this geometric method, we can easily prove the Fermat's Last Theorem.

2. Proof

2.1. Geometric Construction

A triangle has three sides, a, b and c, respectively. Firstly, let us discuss an easy condition: $a = b$.

As we can see in **Figure 1**, the point O is the center of circle, the radius of the circle is r, the point A and B are on the circle, and A is fixed, B can move to B' (B' is B1, B2, …, shown on the circle), connect point O, A and B to form a triangle ΔOAB. The length of each side of ΔOAB is: OA = a, OB = b, AB = c, and $a = b = r$, the angle of $\angle \mathrm{BOB}' = \theta$.

If $\angle \mathrm{BOB}' = \theta = 0$ (B' = B), then $\mathrm{AB} = c = a + b = 2r$, a, b and c satisfy the function $a^n + b^n = c^n$ (2), in this condition, $n = 1$.

If $\angle \mathrm{BOB}' = \theta = \frac{\pi}{2}$ (B' = B3), according to Pythagoras Theorem:

$\mathrm{AB}^2 = c^2 = a^2 + b^2$, a, b and c satisfy the function $a^n + b^n = c^n$ (1), in this condition, $n = 2$.

In general condition: $\angle \mathrm{BOB}' = \theta$, because of $\mathrm{BB}' \perp \mathrm{AB}'$, then $c = 2r\cos\frac{\theta}{2}$, if a, b and c satisfy the function $a^n + b^n = c^n$ (1), then

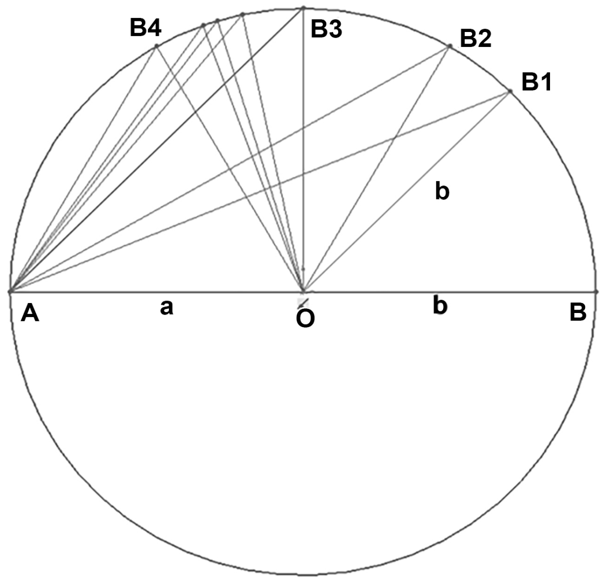


Figure 1. The Geometric construction to prove Fermat's Last Theorem ($a = b$).

$$2r^n = \left(2r\cos\frac{\theta}{2}\right)^n$$

$$2^{\frac{1}{n}} = 2\cos\frac{\theta}{2}$$

$$\frac{1}{n} = \log_2\left(2\cos\frac{\theta}{2}\right) = \log_2\left(\frac{c}{r}\right) \quad (3)$$

The function (3) shows the relationship between n and θ, function (3) is the necessary and sufficient conditions of function (1). We can draw the function (3) as **Figure 2**. ($\theta \in [0, \pi]$).

1) When $\theta = \frac{2\pi}{3} - \epsilon$, ($\epsilon > 0$ and $\epsilon \to 0$),

$$2\cos\frac{\theta}{2} = 2\cos\left(\frac{\pi}{3} - \frac{\epsilon}{2}\right) = 2\left(\cos\frac{\pi}{3} * \cos\frac{\epsilon}{2} + \sin\frac{\pi}{3} * \sin\frac{\epsilon}{2}\right)$$, then

$$= \cos\frac{\epsilon}{2} + \sqrt{3}\sin\frac{\epsilon}{2} \to 1 + \frac{\sqrt{3}}{2}\epsilon$$

$$\frac{1}{n} = \log_2 2\cos\frac{\theta}{2} \to \log_2\left(1 + \frac{\sqrt{3}}{2}\epsilon\right),\ \frac{1}{n} \to \frac{\sqrt{3}}{2\ln 2}\epsilon \to 0$$, and $\frac{1}{n} > 0, n \to +\infty$

2) When $\theta = \frac{2\pi}{3} + \epsilon$, ($\epsilon > 0$ and $\epsilon \to 0$),

$$2\cos\frac{\theta}{2} = 2\cos\left(\frac{\pi}{3} + \frac{\epsilon}{2}\right) = 2\left(\cos\frac{\pi}{3} * \cos\frac{\epsilon}{2} - \sin\frac{\pi}{3} * \sin\frac{\epsilon}{2}\right)$$, then

$$= \cos\frac{\epsilon}{2} - \sqrt{3}\sin\frac{\epsilon}{2} \to 1 - \frac{\sqrt{3}}{2}\epsilon$$

$$\frac{1}{n} = \log_2 2\cos\frac{\theta}{2} \to \log_2\left(1 - \frac{\sqrt{3}}{2}\epsilon\right),\ \frac{1}{n} \to \left(-\frac{\sqrt{3}}{2\ln 2}\epsilon\right) \to 0$$, and $\frac{1}{n} < 0, n \to -\infty$

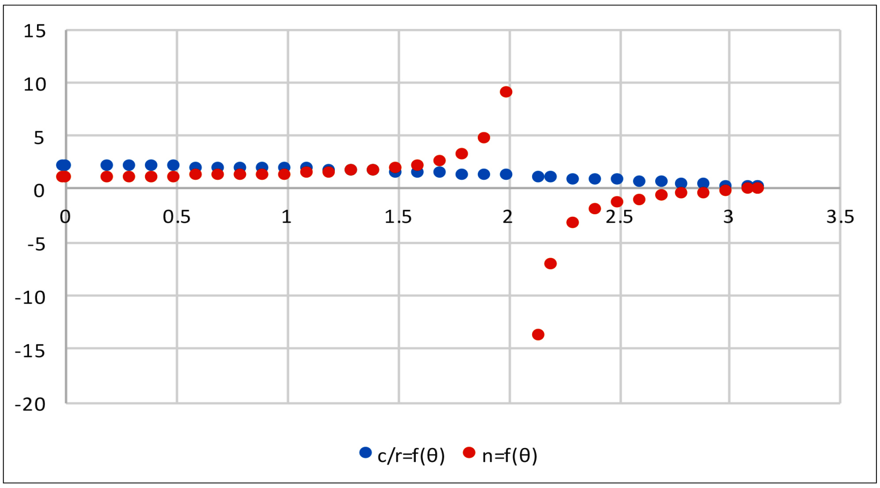


Figure 2. The relationship between n, θ and c ($a = b$).

3) When $\theta = \pi - \epsilon$, ($\epsilon > 0$ and $\epsilon \to 0$),

$$2\cos\frac{\theta}{2} = 2\cos\left(\frac{\pi}{2} - \frac{\epsilon}{2}\right) = 2\left(\cos\frac{\pi}{2} * \cos\frac{\epsilon}{2} + \sin\frac{\pi}{2} * \sin\frac{\epsilon}{2}\right)$$, then

$$= 0 + \sin\frac{\epsilon}{2} \to +\frac{\epsilon}{2}$$

$$\frac{1}{n} = \log_2 2\cos\frac{\theta}{2} \to \log_2\left(\frac{\epsilon}{2}\right) \to -\infty, \quad n \to -0$$

If $a = b = r$ is integer, $n \in \mathbb{Q}$, then $n = \frac{p}{q} > 1$, $(p,q) = 1$, p and q are positive integer.

$$2^{\frac{1}{n}} = 2^{\frac{q}{p}} = 2\cos\frac{\theta}{2} = \frac{c}{r} \tag{3}$$

If c can be a positive integer, $2^{\frac{1}{n}}$ must be a rational number. If $2^{\frac{1}{n}} = 2^{\frac{q}{p}}$ is rational number, then $2^{\frac{q}{p}} = \frac{t}{s}$, $(t,s) = 1$, t and s are positive integer, $0 < \frac{t}{s} < 2$.

$$\left(2^{\frac{q}{p}}\right)^{\frac{p}{q}} = \left(\frac{t}{s}\right)^{\frac{p}{q}} = 2 \tag{4}$$

$$(t)^{\frac{p}{q}} = 2(s)^{\frac{p}{q}}$$

$$\left((t)^{\frac{p}{q}}\right)^q = \left(2(s)^{\frac{p}{q}}\right)^q$$

$$(t)^p = 2^q (s)^p \tag{5}$$

t, s, p, q are positive integer, so 2^q is an even number, then t is an even number, let $t = 2k$ (k is positive integer),

$$(2k)^p = 2^q (s)^p$$

$$2^{(p-q)} k^p = s^p \tag{6}$$

$n = \frac{p}{q} > 1$, $p > q$, *p*, *q* are positive integer, then $(p-q) > 0$ and $(p-q)$ is positive integer, so 2^{p-q} is even number, and $(s)^p$ is also an even number, then *s* must be an even number.

Above all, we can prove that *t* and *s* are even number, which is contradictory with $(t,s) = 1$, therefore, $2^{\frac{1}{n}} = 2^{\frac{q}{p}} = \frac{c}{r}$ is irrational number, *r* is integer, then *c* must be irrational number.

In summary, a, *b* and *c* are the 3 sides of triangle, if $a = b = r$ is positive integer, *n* is rational number ($n \geq 1$), and a, *b*, *c* satisfy the function $a^n + b^n = c^n$ (1), only if *n* = 1, *c* can be integer, the relationship between *n* and *θ* is:

$$\frac{1}{n} = \log_2\left(2\cos\frac{\theta}{2}\right) = \log_2\left(\frac{c}{r}\right)$$

2.2. The Proof of Fermat's Last Theorem

The Fermat's Last Theorem is:

$$a^n + b^n = c^n \tag{1}$$

When *n* is integer and $n > 2$, the function (1) has no positive integer solution, which means a, *b* and *c* can't be positive integer at the same time or when *a*, *b*, and *c* is positive integer, *n* is integer and $n > 2$, the function $a^n + b^n = c^n$ (1) cannot be established.

First of all, we have to prove the value of *a*, *b*, *c* in the function (1) can form a triangle.

If n is integer and $n > 2$, *a*, *b* and *c* are more than 0, then:

$a^n + b^n = c^n > a^n$, so $c > a$;

$a^n + b^n = c^n > b^n$, so $c > b$;

$a^n + b^n = c^n < (a+b)^n$, so $c < a + b$

Therefore, *a*, *b*, and *c* can certainly form a triangle ΔOAB [7], and the triangle ΔOAB is shown in **Figure 3**.

In the section of 1.1, we have proved that the value of *c* in the function (1) is irrational even if *n* is rational ($n > 1$) when $a = b = r$ and *r* is positive integer. To prove the Fermat's Last Theorem, we have to prove another condition: if *a*, *b* and *c* is positive integer, and $a \neq b$, the function $a^n + b^n = c^n$ (2) can not be established when *n* is integer and $n > 2$.

Similarly, we can also construct the geometric method as same as the section of 2.1. The geometric graph is shown in **Figure 3**. As shown in **Figure 3**, OA = a, OB = *b*, and let a > *b*, the point of B can move to B' on the circle which centered on point O and the radius of the circle is *b*. Connect point O, A and B to form a triangle ΔOAB, the angle of $\angle BOB' = \theta$, AB'= *c*. a, *b* and *c* are the length of the 3 sides of ΔOAB, so their value are more than 0.

So $AB'^2 = c^2 = (a + b\cos\theta)^2 + (b\sin\theta)^2 = a^2 + b^2 + 2ab\cos\theta$

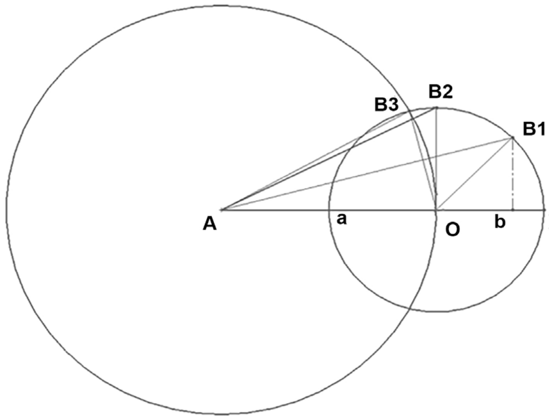


Figure 3. The Geometric construction to prove Fermat's Last Theorem ($a \neq b$ and let $a > b$).

If $\theta = 0$, $c^2 = a^2 + b^2 + 2ab\cos\theta = a^2 + b^2 + 2ab = (a+b)^2$, $c = a + b$, $n = 1$.

If $\theta = \frac{\pi}{2}$, $c^2 = a^2 + b^2 + 2ab\cos\theta = a^2 + b^2$, n = 2, and the value of a, b, c can be integer.

If $\theta \neq \frac{\pi}{2}$ and $\theta \neq 0$, a and b are positive integer, a, b and c satisfy the function (1), then:

$$a^n + b^n = c^n = \left(a^2 + b^2 + 2ab\cos\theta\right)^{\frac{n}{2}}$$

$$c = \sqrt{a^2 + b^2 + 2ab\cos\theta} = \sqrt[n]{a^n + b^n} \quad (7)$$

Function (7) shows the relationship between a, b, n and θ, what's more, function (7) is the necessary and sufficient conditions of function (1), so:

$$c = \sqrt{a^2 + b^2 + 2ab\cos\theta} = \sqrt{a^2 + b^2 + 2ab - 2ab + 2ab\cos\theta}$$

$$c = \sqrt{(a+b)^2 - 2ab(1-\cos\theta)} \quad (8)$$

if c is integer, according to function (8), $c < (a+b)$, let $c = (a+t)$, t is integer and $0 < t < b$.

$$c^2 = (a+t)^2 = (a+b)^2 - 2ab(1-\cos\theta)$$

$$(2a+b+t)(b-t) = 2ab(1-\cos\theta) \quad (9)$$

a, b and t is integer, so $(2a+b+t)(b-t)$ is integer, then $\cos\theta$ must be rational. Let $\cos\theta = R$, and R is rational.

1) When $\theta \in \left(0, \frac{\pi}{2}\right)$**,** $R \in (1,0)$

$$\left(a^n + b^n\right)^2 = a^{2n} + 2a^n b^n + b^{2n} \quad (10)$$

$$\begin{aligned}\left(c^n\right)^2 &= \left(a^2 + b^2 + 2ab\cos\theta\right)^n > \left(a^2 + b^2\right)^n \\ &= a^{2n} + C_n^1 a^{2(n-1)} b^{2\times 1} + C_n^2 a^{2(n-2)} b^{2\times 2} + \cdots + C_n^n a^{2(n-n)} b^{2\times n}\end{aligned} \quad (11)$$

So $a^{2n} + 2a^n b^n + b^{2n} < \left(a^2 + b^2\right)^n < c^{2n}$, $a^n + b^n < c^n$ (12).

Therefore, when $\theta \in \left(0, \frac{\pi}{2}\right)$, $R \in (1,0)$, $c = \sqrt{a^2 + 2Rab + b^2}$ and c can be integer, n is integer and n>2, if a, b is positive integer, the relationship between a, b, c and n must be $a^n + b^n < c^n$, no matter c is integer or not.

2) When $\theta \in \left(\frac{\pi}{2}, \pi\right]$, $R \in (0,-1)$

Because of $a^n + b^n = c^n$, n is integer and $n > 2$, so that $c > a$ and $c > b$. However, if B' = B_3 (B_3 is the intersection of the two circles in **Figure 3**), $\theta = \angle BOB_3 = \frac{\pi}{2} + \arcsin\frac{b}{2a}$ (please see **Figure 3**.), it means $AB' = AB_3 = AO = c = a$, and if $\theta = \angle BOB' > \angle BOB_3$, $AB' = c < a$ (According to Euclid's Elements, big angle to big side) [7], so if $\theta \in \left(\frac{\pi}{2} + \arcsin\frac{b}{2a}, \pi\right]$, the value of $c < a$.

Therefore, if $\theta \in \left(\frac{\pi}{2}, \frac{\pi}{2} + \arcsin\frac{b}{2a}\right)$, $R \in \left(-\frac{b}{2a}, 0\right)$, then $c > a$, and there is possible to make the function $a^n + b^n = c^n$ to be tenable when $n > 2$. Let $c = (a+t)$, t is integer and $0 < t < b$, $a \geq b+1$, then

$$a^n + b^n = c^n = (a+t)^n \geq (a+1)^n \tag{13}$$

$$(a+1)^n = a^n + C_n^1 a^{(n-1)} + C_n^2 a^{(n-2)} + \cdots + 1 \tag{14}$$

So, if the function $a^n + b^n = c^n$ is established when n is integer and $n > 2$, $a > b$, we can get:

$$a^n + b^n \geq a^n + C_n^1 a^{(n-1)} + C_n^2 a^{(n-2)} + \cdots + C_n^{(n-1)} a^1 + 1 \tag{15}$$

$$1 \geq \frac{C_n^1 a^{(n-1)} + C_n^2 a^{(n-2)} + \cdots + 1}{b^n} \geq \frac{C_n^1 (b+1)^{(n-1)} + C_n^2 (b+1)^{(n-2)} + \cdots + 1}{b^n} \tag{16}$$

If $b = 1$, according to function (15), $1 \geq C_n^1 a^{(n-1)} + C_n^2 a^{(n-2)} + \cdots + 1$ can't be established; if $b > 1$, according to function (16), there are two conditions that need to be discussed:

1) $n \geq b+1$, then

$$\frac{C_n^1 (b+1)^{(n-1)} + C_n^2 (b+1)^{(n-2)} + \cdots + 1}{b^n}$$
$$> \frac{(b+1)(b+1)^{(n-1)} + (b+1)(b+1)^{(n-2)} + \cdots + (b+1)(b+1)^1 + b + 1 + 1 - b - 1}{b^n}$$
$$= \left(\frac{(b+1)^{n+1} - 1}{(b+1) - 1} - (b+1)\right) \times \frac{1}{b^n}$$
$$= \frac{(b+1)^{n+1} - 1 - b(b+1)}{b^{n+1}} > 1$$

(which is contradictory with function (16))

Therefore, if $b > 1$ and $n \geq b+1$, the relationship between a, b, c and n must be $a^n + b^n < c^n$, the function (16) can't be satisfied.

2) $2 < n < b+1$, then

$$a^n + b^n = c^n = (a+t)^n \tag{17}$$

$$a^n + b^n = a^n + C_n^1 a^{(n-1)} t + C_n^2 a^{(n-2)} t^2 + \cdots + t^n$$

$$t^n + C_n^{(n-1)} a^1 t^{(n-1)} + \cdots + C_n^2 a^{(n-2)} t^2 + C_n^1 a^{(n-1)} t - b^n = 0 \tag{18}$$

If there exists a positive integer c to satisfy the function (17), then t must be a integer, and the value of t is the solution for the equation of degree n with one unknown in function (18), and the solution of t in function (18) is also the solution of t in function (17).

Obviously, one of the solution of t in function (17) must be:

$$t = -a \pm \sqrt[n]{a^n + b^n} \ (\text{if } n \text{ is even number}) \tag{19}$$

$$t = -a + \sqrt[n]{a^n + b^n} \ (\text{if } n \text{ is odd number}) \tag{20}$$

In this paper, we only discuss the condition of $0 < t < b$, so the value of t satisfies the requirement is:

$$t = -a + \sqrt[n]{a^n + b^n} \tag{21}$$

Therefore, we have found a solution for the equation of degree n with one unknown in function (18), and the solution of t is the function (19) and (20).

For example: let a = 2, b = 1, and n = 5, so the function (18) is equal to the function shows below:

$$t^5 + 10t^4 + 40t^3 + 80t^2 + 80t - 1 = 0 \tag{22}$$

According to function (20), we can easily get the solution of t in function (22) is

$$t = -2 + \sqrt[5]{2^5 + 1^5} = -2 + \sqrt[5]{33} \tag{23}$$

This is a special example of the solution for the equation of degree n with one unknown. However, how to find the solution for the arbitrarily equation of degree n with one unknown as function (24):

$$A_n t^n + A_{(n-1)} t^{(n-1)} + \cdots + A_2 t^2 + A_1 t - B = 0 \tag{24}$$

Let's discuss the question in the next Section 2.3.

According to function (18), we can transform it to the function (25):

$$t^{n-1} + C_n^{(n-1)} a^1 t^{(n-2)} + \cdots + C_n^2 a^{(n-2)} t^1 + C_n^1 a^{(n-1)} = \frac{b^n}{t} \tag{25}$$

t is a positive integer, so the left side of the equal sign in function (25) must be a positive integer, therefore, the value of $\frac{b^n}{t}$ must be a positive integer.

Let the value of $\frac{b^n}{t} = m$, and m is a positive integer, then

$$t^{n-1} + C_n^{(n-1)} a^1 t^{(n-2)} + \cdots + C_n^2 a^{(n-2)} t^1 + C_n^1 a^{(n-1)} - m = 0 \tag{26}$$

Therefore, we have to find the value of t in the function (26), and determine if the value of t can be an integer. If t can't be a positive integer when $n \ge 3$, then the Fermat's last Theorem is right.

If $n = 1$, the function (26) is equal to $1-m=0$, then $\frac{b^1}{t}=1$, $t=b$, so $c=a+t=a+b$;

If $n = 2$, the function (26) is equal to $t+2a-m=0$, then $\frac{b^2}{t}=t+2a$, $t^2+2at=b^2$, so $a^2+b^2=a^2+t^2+2at=(a+t)^2=c^2$.

Therefore, if $n = 1$ and $n = 2$, the function $a^n+b^n=c^n$ can have integer solution.

If $n = 3$, the function (26) is equal to $t^2+3at+3a^2-m=0$, then

$t=\frac{-3a\pm\sqrt{4m-3a^2}}{2}$,

The value of t must be integer and $0<t<b$, then $4m-3a^2=R^2a^2$, R^2 is a positive integer, thus, $4m=Ka^2$, $K-3=R^2$, $K=4,7,12,19,\cdots$, $R=1,2,3,4,\cdots$, then:

$$t=\frac{-3a+\sqrt{4m-3a^2}}{2}=\frac{(R-3)a}{2}$$

$$\frac{b^3}{t}=m=\frac{Ka^2}{4}=\frac{(R^2+3)a^2}{4}$$

$$b^3=\frac{(R^2+3)a^2}{4}\times\frac{(R-3)a}{2}=\frac{(R^2+3)(R-3)a^3}{8}$$

$$b=\frac{a}{2}\sqrt[3]{(R^2+3)(R-3)}$$

Therefore, $(R^2+3)(R-3)$ must be a positive cubic number, so $R>3$. Let $(R^2+3)(R-3)=(R-\alpha)^3$, α is positive integer, then:

$$R^3-3R^2+3R-9=R^3-3\alpha R^2+3\alpha^2R-\alpha^3$$

If $\alpha=1$, $R^3-3\alpha R^2+3\alpha^2R-\alpha^3=R^3-3R^2+3R-1>R^3-3R^2+3R-9$

If $\alpha=2$, $\begin{aligned}&R^3-3\alpha R^2+3\alpha^2R-\alpha^3-(R^3-3R^2+3R-9)\\&=-3R^2+9R+1=\Delta=-3\left(R-\frac{3}{2}\right)^2+\frac{31}{4}\end{aligned}$

Because R is positive integer and $R>3,\Delta<0$, so, $R^3-6R^2+12R-8<R^3-3R^2+3R-9$

If $\alpha=3$, $\begin{aligned}&R^3-3\alpha R^2+3\alpha^2R-\alpha^3-(R^3-3R^2+3R-9)\\&=-6R^2+24R-18=\Delta=-6\left[(R-2)^2-1\right]\end{aligned}$, Because R is positive integer and $R>3,\Delta<0$, so, $R^3-9R^2+27R-27<R^3-3R^2+3R-9$

If $\alpha=\alpha$, and $\alpha>3$, then,

$$\begin{aligned}&R^3-3\alpha R^2+3\alpha^2R-\alpha^3-(R^3-3R^2+3R-9)\\&=-(3\alpha-3)R^2+(3\alpha^2-3)R-(\alpha^3-9)\\&=-(3\alpha-3)\left[R^2-(\alpha+1)R\right]-(\alpha^3-9)\\&=\Delta=-(3\alpha-3)\left[R-\left(\frac{\alpha+1}{2}\right)\right]^2+\frac{-\alpha^3+3\alpha^2-3\alpha+33}{4}\end{aligned}$$

When $\alpha = 4$, $\frac{-\alpha^3 + 3\alpha^2 - 3\alpha + 33}{4} = \frac{5}{4}$, Because R is positive integer and $R > 3, \Delta < 0$, so, $R^3 - 3\alpha R^2 + 3\alpha^2 R - \alpha^3 < R^3 - 3R^2 + 3R - 9$

When $\alpha = 5$, $\frac{-\alpha^3 + 3\alpha^2 - 3\alpha + 33}{4} = -8$, No matter what the value of R is, $R^3 - 3\alpha R^2 + 3\alpha^2 R - \alpha^3 < R^3 - 3R^2 + 3R - 9$

The function $f(\alpha) = -\alpha^3 + 3\alpha^2 - 3\alpha + 33$ is monotonically decreasing function when $\alpha > 5$, so $f(\alpha) < -8$ when $\alpha > 5$, therefore, No matter what the value of R is, $R^3 - 3\alpha R^2 + 3\alpha^2 R - \alpha^3 < R^3 - 3R^2 + 3R - 9$.

Thus, $(R^2 + 3)(R - 3) = (R - \alpha)^3$ can't be established when α is positive integer, and $(R^2 + 3)(R - 3)$ can't be a positive cubic number, therefore, the value of b can't be a positive integer.

In conclusion, if $n = 3$, and the solution of t in function (26) is positive integer, but the value of b can't be a positive integer, thus, the Fermat's Last Theorem is established when $n = 3$.

2.3. The Solution for the Equation of Degree *n* with One Unknown

How to find the solution for the arbitrarily equation of degree n with one unknown as function (24):

$$A_n t^n + A_{(n-1)} t^{(n-1)} + \cdots + A_2 t^2 + A_1 t - B = 0 \tag{24}$$

This paper only discusses the real solution of function (24). As we already known, the solution of t in the function (18) is

$$t = -a \pm \sqrt[n]{a^n + b^n} \; (\text{if } n \text{ is even number}) \tag{19}$$

$$t = -a + \sqrt[n]{a^n + b^n} \; (\text{if } n \text{ is odd number}) \tag{20}$$

So, the function (24) can be deformed as function (27):

$$\begin{aligned} &A_n\left(t^n + C_n^{(n-1)} a^1 t^{(n-1)} + \cdots + C_n^2 a^{(n-2)} t^2 + C_n^1 a^{(n-1)} t - b_n\right) \\ &+B_{(n-1)}\left(t^{(n-1)} + C_{(n-1)}^{(n-2)} a^1 t^{(n-2)} + \cdots + C_{(n-1)}^2 a^{(n-2)} t^2 + C_{(n-1)}^1 a^{(n-1)} t - b_{(n-1)}\right) \\ &+B_{(n-2)}\left(t^{(n-2)} + C_{(n-2)}^{(n-3)} a^1 t^{(n-3)} + \cdots + C_{(n-2)}^2 a^{(n-4)} t^2 + C_{(n-2)}^1 a^{(n-3)} t - b_{(n-2)}\right) \\ &+\cdots + B_2\left(t^2 + C_2^1 a^1 t^1 - b_2\right) + B_1\left(t - b_1\right) = 0 \end{aligned} \tag{27}$$

The coefficients in function (27) are shown below:

$$B_{(n-1)} = \left(A_{(n-1)} - A_n C_n^{(n-1)} a^1\right)$$

$$\begin{aligned} B_{(n-2)} &= \left[A_{(n-2)} - A_n C_n^{(n-2)} a^2 - B_{(n-1)} C_{(n-1)}^{(n-2)} a^1\right] \\ &= \left[A_{(n-2)} - A_n C_n^{(n-2)} a^2 - \left(A_{(n-1)} - A_n C_n^{(n-1)} a^1\right) C_{(n-1)}^{(n-2)} a^1\right] \end{aligned}$$

$$\vdots$$

$$B_2 = \left[A_2 - A_n C_n^2 a^{(n-2)} - B_{(n-1)} C_{(n-1)}^2 a^{(n-2)} - B_{(n-2)} C_{(n-2)}^2 a^{(n-4)} - \cdots - B_3 C_3^2 a^1\right]$$

$$B_1 = \left[A_1 - A_n C_n^1 a^{(n-1)} - B_{(n-1)} C_{(n-1)}^1 a^{(n-1)} - B_{(n-2)} C_{(n-2)}^1 a^{(n-3)} - \cdots - B_2 C_2^1 a^1\right]$$

$$A_n b_n + B_{(n-1)} b_{(n-1)} + B_{(n-2)} b_{(n-2)} + \cdots + B_2 b_2 + B_1 b_1 = B \tag{28}$$

The solution of t in the function (27) is $t(k)$, and the value $t(k)$ after substituting the equation coefficient with $A_n, B_{(n-1)}, B_{(n-2)}, \cdots, B_2$ and B_1 in the function (27) is zero. According to the function (19) and (20), we can easily get the value $t(k)\,(k = 1,2,\cdots,n)$ are:

As $0 < t < b$, so:

$$t(n) = -a + \sqrt[n]{a^n + b_n}$$

$$t(n-1) = -a + \sqrt[(n-1)]{a^{(n-1)} + b_{(n-1)}}$$

$$t(n-2) = -a + \sqrt[(n-2)]{a^{(n-2)} + b_{(n-2)}}$$

$$\vdots$$

$$t(2) = -a + \sqrt[2]{a^2 + b_2}$$

$$t(1) = b_1$$

If we want the function (27) to be established, then all the value of $t(k)$ must be equal, and we can find the solution for the equation of degree n with one unknown. We can set the value of a and $a \neq 0$, so here are n equations and we have to find the solution of $b_1, b_2, \cdots, b_n$, and the solution can be found.

1) Part one: let's find the solution for the random equation of degree 3 with one unknown, the equation is shown below:

$$A_3 t^3 + A_2 t^2 + A_1 t - B = 0 \tag{29}$$

The function (29) can be deformed as:

$$A_3\left(t^3 + 3at^2 + 3a^2 t - b_3\right) + \left(A_1 - A_3 \times 3a^2\right)t - \left(B - A_3 b_3\right) = 0 \tag{30}$$

and $A_3 \times 3a = A_2$, $a = \dfrac{A_2}{3A_3}$

we can easily get the value $t(3)$ and $t(1)$:

$$\begin{cases} t(3) = -\dfrac{A_2}{3A_3} + \sqrt[3]{\left(\dfrac{A_2}{3A_3}\right)^3 + b_3} \\ t(1) = \dfrac{B - A_3 b_3}{A_1 - \dfrac{A_2^2}{3A_3}} \end{cases}$$

In order to make the function (29) to be established, the value of $t(3)$ and $t(1)$ must be equal, let $t(1) = t(3) = m$, then:

$$b_3 = \frac{B - m\left(A_1 - \dfrac{A_2^2}{3A_3}\right)}{A_3}$$

$$\left(m+\frac{A_2}{3A_3}\right)^3=\left(\frac{A_2}{3A_3}\right)^3+b_3=\left(\frac{A_2}{3A_3}\right)^3+\frac{B-m\left(A_1-\frac{A_2^2}{3A_3}\right)}{A_3}$$

$$\left(m+\frac{A_2}{3A_3}\right)^3=-\frac{A_1-\frac{A_2^2}{3A_3}}{A_3}\left(m+\frac{A_2}{3A_3}\right)+\frac{27A_3^2B+9A_1A_2A_3-2A_2^3}{27A_3^3}$$

Let $m+\frac{A_2}{3A_3}=w$, then:

$$w^3+\frac{3A_1A_3-A_2^2}{3A_3^2}w-\frac{27A_3^2B+9A_1A_2A_3-2A_2^3}{27A_3^3}=0 \tag{31}$$

According to the Cardano Formula of the General Solution of Cubic Algebraic Equations [8], the real solution of function (31) is:

$$w=\sqrt[3]{-\frac{q}{2}+\sqrt{\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3}}+\sqrt[3]{-\frac{q}{2}-\sqrt{\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3}} \tag{32}$$

$$\begin{cases} p=\frac{3A_1A_3-A_2^2}{3A_3^2} \\ q=-\frac{27A_3^2B+9A_1A_2A_3-2A_2^3}{27A_3^3} \end{cases}$$

Therefore, the solution of t satisfy the function in (29) is $t=-\frac{A_2}{3A_3}+w$

For example, Let's find the solution of t in the function (33) below:

$$2t^3+3t^2+4t-5=0 \tag{33}$$

$$p=\frac{3A_1A_3-A_2^2}{3A_3^2}=\frac{5}{4}$$

$$q=-\frac{27A_3^2B+9A_1A_2A_3-2A_2^3}{27A_3^3}=-\frac{13}{4}$$

$$w=\sqrt[3]{\frac{13}{8}+\sqrt{\left(\frac{13}{8}\right)^2+\left(\frac{5}{12}\right)^3}}+\sqrt[3]{\frac{13}{8}-\sqrt{\left(\frac{13}{8}\right)^2+\left(\frac{5}{12}\right)^3}}\approx 1.20393969$$

so the real solution of t in the function (33) is $t=-\frac{1}{2}+w\approx 0.70393969$

If $n = 4$, the function (26) is equal to the function (34) shows below:

$$t^3+4a^1t^2+6a^2t^1-\left(m-4a^3\right)=0 \tag{34}$$

$$p=\frac{3A_1A_3-A_2^2}{3A_3^2}=\frac{2a^2}{3}$$

$$q=-\frac{27A_3^2B+9A_1A_2A_3-2A_2^3}{27A_3^3}=\frac{20a^3}{27}-m$$

$$w=\sqrt[3]{-\frac{q}{2}+\sqrt{\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3}}+\sqrt[3]{-\frac{q}{2}-\sqrt{\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3}}$$

$$u=\sqrt{\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3}=\sqrt{\left(\frac{10a^3}{27}-\frac{m}{2}\right)^2+\left(\frac{2a^2}{9}\right)^3} \quad (35)$$

The value of u in the function (35) must be a integer, therefore, $\left(\frac{10a^3}{27}-\frac{m}{2}\right)^2+\left(\frac{2a^2}{9}\right)^3$ must be a positive square integer, otherwise, the value of w can't be a rational number.

Let $m = 2k$, $a = 3s$, k and s is positive integer, then:

$$u=\sqrt{\left(\frac{10a^3}{27}-\frac{m}{2}\right)^2+\left(\frac{2a^2}{9}\right)^3}=\sqrt{\left(10s^3-k\right)^2+\left(2s^2\right)^3}$$

Obviously, if $\left(10s^3-k\right)^2=K^2s^6$, $K^2+8=R^2$, K and R is a positive integer, under this condition, the value of u can be a integer. Then, $K^2=1,8,17,28,\cdots$, $R=3,4,5,6,\cdots$, and $u=\pm Rs^3$. However, the value of K must be a positive square integer, a positive square integer plus 8 is still a square integer, only when $K^2=1, R^2=9$ can satisfy the requirement, then $k=9s^3$ or $11s^3$, $u=\pm 3s^3$.

$$p=\frac{2a^2}{3}=6s^2$$

$$q=-\frac{27A_3^2B+9A_1A_2A_3-2A_2^3}{27A_3^3}=\frac{20a^3}{27}-m=20s^3-2k=\pm 2s^3$$

When $k=9s^3$

$$w=\sqrt[3]{s^3+\left(\pm 3s^3\right)}+\sqrt[3]{s^3-\left(\pm 3s^3\right)}=\sqrt[3]{4}s+\sqrt[3]{-2}s$$

When $k=11s^3$

$$w=\sqrt[3]{-s^3+\left(\pm 3s^3\right)}+\sqrt[3]{-s^3-\left(\pm 3s^3\right)}=\sqrt[3]{2}s+\sqrt[3]{-4}s$$

Absolutely, the value of w is irrational, therefore, the solution of t satisfy the function in (34) is $t=-12s+w$ is irrational.

In conclusion, if $n = 4$, the solution of t in function (26) is irrational, thus, the Fermat's Last Theorem is established when $n = 4$.

2) Part two: let's discuss the solution for the random equation of degree 4 with one unknown, the equation is shown below:

$$A_4t^4+A_3t^3+A_2t^2+A_1t-B=0 \quad (32)$$

The function (32) can be deformed as:

$$A_4\left(t^4+4at^3+6a^2t^2+4a^3t-b_4\right)+\left(A_2-A_4\times 6a^2\right)t^2 + \left(A_1-A_4\times 4a^3\right)t-\left(B-A_4b_4\right)=0 \quad (33)$$

and $A_4\times 4a=A_3$, $a=\frac{A_3}{4A_4}$

we can easily get the value $t(4)$ and $t(2)$:

$$\begin{cases} t(4) = -\dfrac{A_3}{4A_4} + \sqrt[4]{\left(\dfrac{A_3}{4A_4}\right)^4 + b_4} \\ t(2) = \dfrac{-\left(A_1 - \dfrac{A_3^3}{16A_4^2}\right) \pm \sqrt{\left(A_1 - \dfrac{A_3^3}{16A_4^2}\right)^2 + 4\left(A_2 - \dfrac{3A_3^2}{8A_4}\right)(B - A_4 b_4)}}{2\left(A_2 - \dfrac{3A_3^2}{8A_4}\right)} \end{cases}$$

Similarly, in order to make the function (33) to be established, the value of $t(4)$ and $t(2)$ must be equal, let $t(4) = t(2) = m$, then

$$b_4 = \frac{B - \left(A_2 - \dfrac{3A_3^2}{8A_4}\right)m^2 - \left(A_1 - \dfrac{A_3^3}{16A_4^2}\right)m}{A_4}$$

$$\left(m + \frac{A_3}{4A_4}\right)^4 = \left(\frac{A_3}{4A_4}\right)^4 + b_4$$

$$= \left(\frac{A_3}{4A_4}\right)^4 + \frac{B - \left(A_2 - \dfrac{3A_3^2}{8A_4}\right)m^2 - \left(A_1 - \dfrac{A_3^3}{16A_4^2}\right)m}{A_4}$$

$$\left(m + \frac{A_3}{4A_4}\right)^4 + \frac{8A_2A_4 - 3A_3^2}{8A_4^2}\left(m + \frac{A_3}{4A_4}\right)^2$$

$$-\frac{8A_2A_3A_4^2 - 3A_3^3A_4 - 16A_1A_4^2 + A_3^3}{16A_4^3}\left(m + \frac{A_3}{4A_4}\right)$$

$$= \frac{128A_4^3B + 8A_2A_3^2A_4 - 16A_2A_3^2A_4^2 + 6A_3^4A_4 + 32A_1A_3A_4^2 - \dfrac{9}{2}A_3^4}{128A_4^4}$$

Let $m + \dfrac{A_3}{4A_4} = w$, then:

$$w^4 + pw^2 + qw + r = 0 \tag{34}$$

$$\begin{cases} p = \dfrac{8A_2A_4 - 3A_3^2}{8A_4^2} \\ q = -\dfrac{8A_2A_3A_4^2 - 3A_3^3A_4 - 16A_1A_4^2 + A_3^3}{16A_4^3} \\ r = -\dfrac{128A_4^3B + 8A_2A_3^2A_4 - 16A_2A_3^2A_4^2 + 6A_3^4A_4 + 32A_1A_3A_4^2 - \dfrac{9}{2}A_3^4}{128A_4^4} \end{cases}$$

then the function (34) can be transformed to the function (35):

$$w^4 + pw^2 + qw + r = \left(w^2 + kw + u\right)\left(w^2 - kw + v\right) = 0 \tag{35}$$

$$\left(w^2 + kw + u\right)\left(w^2 - kw + v\right) = w^4 + \left(u + v - k^2\right)w^2 + k\left(v - u\right)w + uv$$

Then we can get:

$$\begin{cases} p = u + v - k^2 \\ q = k(v - u) \\ r = uv \end{cases}$$

We can set a value for k and $k \neq 0$, p, q, and r are the known number, therefore, we can find the value of u and v are:

$$u = \frac{k^3 + pk - q}{2k}$$

$$v = \frac{k^3 + pk + q}{2k}$$

$$\frac{k^3 + pk - q}{2k} \times \frac{k^3 + pk + q}{2k} = r$$

$$k^6 + 2pk^4 + (p^2 - 4r)k^2 - q^2 = 0 \quad (36)$$

Let $k^2 = x$, then the function (36) can be deformed as :

$$x^3 + 2px^2 + (p^2 - 4r)x - q^2 = 0 \quad (37)$$

Now, the problem is transformed to find the solution x for function (37), according to the results of **Part one**, we can get the solution x is:

$$W = \sqrt[3]{-\frac{Q}{2} + \sqrt{\left(\frac{Q}{2}\right)^2 + \left(\frac{P}{3}\right)^3}} + \sqrt[3]{-\frac{Q}{2} - \sqrt{\left(\frac{Q}{2}\right)^2 + \left(\frac{P}{3}\right)^3}} \quad (38)$$

$$P = \frac{3(p^2 - 4r) - 4p^2}{3} = -\frac{p^2 + 12r}{3}$$

$$Q = -\frac{27q^2 + 18(p^2 - 4r)p - 16p^3}{27} = -\frac{27q^2 - 72pr + 2p^3}{27}$$

And the solution $x = -\frac{2p}{3} + W$

$$k_0 = k_{1,2} = \pm\sqrt{x} = \sqrt{-\frac{2p}{3} + W}$$

$$u_0 = \frac{k_0^3 + pk_0 - q}{2k_0}$$

$$v_0 = \frac{k_0^3 + pk_0 + q}{2k_0}$$

$$(w^2 + k_0 w + u_0)(w^2 - k_0 w + v_0) = 0$$

Then the solutions w in the function (35) are:

$$w_{1,2} = \frac{-k_0 \pm \sqrt{k_0^2 - 4u_0}}{2}$$

$$w_{3,4} = \frac{k_0 \pm \sqrt{k_0^2 - 4v_0}}{2}$$

Therefore, the solution of t in the function (32) are:

$$t_{1,2} = -\frac{A_3}{4A_4} + w_{1,2}$$

$$t_{3,4} = -\frac{A_3}{4A_4} + w_{3,4}$$

For example, Let's find the solution of t in the function (39) below:

$$t^4 + 2t^3 + 3t^2 + 4t - 5 = 0 \tag{39}$$

$$p = \frac{8A_2A_4 - 3A_3^2}{8A_4^2} = \frac{3}{2}$$

$$q = -\frac{8A_2A_3A_4^2 - 3A_3^3A_4 - 16A_1A_4^2 + A_3^3}{16A_4^3} = 2$$

$$r = -\frac{128A_4^3B + 8A_2A_3^2A_4 - 16A_2A_3^2A_4^2 + 6A_3^4A_4 + 32A_1A_3A_4^2 - \frac{9}{2}A_3^4}{128A_4^4} = -\frac{103}{16}$$

$$P = -\frac{p^2 + 12r}{3} = 25$$

$$Q = -\frac{27q^2 - 72pr + 2p^3}{27} = -30$$

$$W = \sqrt[3]{-\frac{Q}{2} + \sqrt{\left(\frac{Q}{2}\right)^2 + \left(\frac{P}{3}\right)^3}} + \sqrt[3]{-\frac{Q}{2} - \sqrt{\left(\frac{Q}{2}\right)^2 + \left(\frac{P}{3}\right)^3}} \approx 1.14063859$$

$$x = -\frac{2p}{3} + W \approx -1 + 1.14063859 = 0.14063859$$

$$k_1 = \sqrt{x} = \sqrt{-\frac{2p}{3} + W} \approx 0.37501813$$

$$u_1 = \frac{k_1^3 + pk_1 - q}{2k_0} \approx -1.8462185$$

$$v_1 = \frac{k_1^3 + pk_1 + q}{2k_0} \approx 3.48685708$$

$$w_{1,2} = \frac{-k_1 \pm \sqrt{k_1^2 - 4u_1}}{2} \approx 1.18412432 \text{ or } -1.5591424$$

$$k_2 = -\sqrt{x} = \sqrt{-\frac{2p}{3} + W} \approx -0.37501813$$

$$u_2 = \frac{k_2^3 + pk_2 - q}{2k_2} \approx 3.48685708$$

$$v_2 = \frac{k_2^3 + pk_2 + q}{2k_2} \approx -1.8462185$$

$$w_{3,4} = \frac{k_2 \pm \sqrt{k_2^2 - 4u_2}}{2} = 0.1875091 + 1.8578744i \text{ or } 0.1875091 - 1.8578744i$$

Therefore, the solutions of t in the function (32) are:

$$t_{1,2} = -\frac{A_3}{4A_4} + w_{1,2} \approx 0.68412432 \text{ or } -2.0591424$$

$$t_{3,4} = -\frac{A_3}{4A_4} + w_{3,4} = -0.3124909 + 1.8578744i \text{ or } -0.3124909 - 1.8578744i$$

Similarly, if $n = 5$, the function (26) is equal to the function (40) shows below:

$$t^4 + 5at^3 + 10a^2t^2 + 10a^3t^1 - \left(m - 5a^4\right) = 0 \tag{40}$$

$$p = \frac{8A_2A_4 - 3A_3^2}{8A_4^2} = \frac{5}{8}a^2$$

$$q = -\frac{8A_2A_3A_4^2 - 3A_3^3A_4 - 16A_1A_4^2 + A_3^3}{16A_4^3} = \frac{5}{8}a^3$$

$$r = -\frac{128A_4^3B + 8A_2A_3^2A_4 - 16A_2A_3^2A_4^2 + 6A_3^4A_4 + 32A_1A_3A_4^2 - \frac{9}{2}A_3^4}{128A_4^4} = \frac{205}{256}a^4 - m$$

$$P = -\frac{p^2 + 12r}{3} = 4m - \frac{10}{3}a^4$$

$$Q = -\frac{27q^2 - 72pr + 2p^3}{27} = -\frac{45ma^2 - 25a^6}{27}$$

$$W = \sqrt[3]{-\frac{Q}{2} + \sqrt{\left(\frac{Q}{2}\right)^2 + \left(\frac{P}{3}\right)^3}} + \sqrt[3]{-\frac{Q}{2} - \sqrt{\left(\frac{Q}{2}\right)^2 + \left(\frac{P}{3}\right)^3}}$$

$$u = \sqrt{\left(\frac{Q}{2}\right)^2 + \left(\frac{P}{3}\right)^3} = \sqrt{\left(\frac{25a^6}{54} - \frac{45ma^2}{54}\right)^2 + \left(\frac{4m}{3} - \frac{10a^4}{9}\right)^3} \tag{41}$$

The value of u in the function (41) must be a integer, therefore, $\left(\frac{25a^6}{54} - \frac{45ma^2}{54}\right)^2 + \left(\frac{4m}{3} - \frac{10a^4}{9}\right)^3$ must be a positive square integer, otherwise, the value of w can't be a rational number.

Let $m = ka^4$, k is positive number, then:

$$u = \sqrt{\left(\frac{25 - 45k}{54}\right)^2 a^{12} + \left(\frac{12k - 10}{9}\right)^3 a^{12}} = \sqrt{\alpha^2 + \beta^3}\,a^6$$

$$-\frac{Q}{2} + u = \left(-\alpha + \sqrt{\alpha^2 + \beta^3}\right)a^6$$

$$W = \sqrt[3]{-\alpha + \sqrt{\alpha^2 + \beta^3}}\,a^2 + \sqrt[3]{-\alpha - \sqrt{\alpha^2 + \beta^3}}\,a^2 = \left(K + S\right)a^2$$

$$x = -\frac{2p}{3} + W = -\frac{5}{12}a^2 + \left(K + S\right)a^2$$

$$k_0 = \pm\sqrt{x} = \pm a\sqrt{\left(K + S\right) - \frac{5}{12}} = \pm\gamma a$$

$$u_0=\frac{k_0^3+pk_0-q}{2k_0}=\frac{(K+S)-\frac{5}{12}a^2+\frac{5}{16}a^2-\frac{5}{16\gamma}a^2}{2}$$

$$v_0=\frac{k_0^3+pk_0+q}{2k_0}$$

$$w_{1,2}=\frac{-k_0\pm\sqrt{k_0^2-4u_0}}{2}$$

$$w_{3,4}=\frac{k_0\pm\sqrt{k_0^2-4u_0}}{2}$$

$$t_{1,2}=-\frac{A_3}{4A_4}+w_{1,2}$$

$$t_{3,4}=-\frac{A_3}{4A_4}+w_{3,4}$$

Obviously,

$$k_0^2-4u_0=\left[(K+S)-\frac{5}{12}\right]a^2-4\left[\frac{(K+S)-\frac{5}{12}a^2+\frac{5}{16}a^2-\frac{5}{16\gamma}a^2}{2}\right]$$

$$=\left[\frac{5}{4\gamma}-\frac{5}{4}-\gamma^2\right]a^2$$

must be a square number, then $\frac{a^2}{4\gamma^2}\left[5\gamma-5\gamma^2-4\gamma^4\right]$ must be a square number,

$$\left[5\gamma-5\gamma^2-4\gamma^4\right]=\delta^2=\left(u\gamma^2-v\right)^2=v^2-2uv\gamma^2+u^2\gamma^4 \qquad (42).$$

Clearly, the function (42) can't be established, therefore, the value of w is irrational, and the solutions of t satisfy the function in (40) is $t=-\frac{5}{4}a+w$ is irrational.

In conclusion, if $n = 5$, the solution of t in function (26) is irrational, thus, the Fermat's Last Theorem is established when $n = 5$.

3) Part three: let's discuss the solution for the random equation of degree 5 with one unknown, the equation is shown below:

$$A_5t^5+A_4t^4+A_3t^3+A_2t^2+A_1t-B=0 \qquad (43)$$

The function (43) can be deformed as:

$$\begin{aligned}&A_5\left(t^5+5at^4+10a^2t^3+10a^3t^2+5a^4t-b_5\right)+\left(A_3-A_5\times10a^2\right)t^3\\&+\left(A_2-A_5\times10a^3\right)t^2+\left(A_1-A_5\times5a^4\right)t-\left(B-A_5b_5\right)=0\end{aligned} \qquad (44)$$

and $A_5\times5a=A_4$, $a=\frac{A_4}{5A_5}$

The function (44) can also be deformed as:

$$A_5\left(t^5+5at^4+10a^2t^3+10a^3t^2+5a^4t-b_5\right)+\left(A_3-A_5\times 10a^2\right)\left(t^3+3ct^2+3c^2t-b_3\right)$$
$$+\left[A_1-A_5\times 5a^4-\left(A_3-A_5\times 10a^2\right)\times 3c^2\right]t-\left[B-A_5b_5-\left(A_3-A_5\times 10a^2\right)b_3\right]=0$$

and $a=\dfrac{A_4}{5A_5}$, $\left(A_3-A_5\times 10a^2\right)\times 3c=\left(A_2-A_5\times 10a^3\right)$, $c=\dfrac{A_2-A_5\times 10a^3}{3\left(A_3-A_5\times 10a^2\right)}$

We can easily get the value $t(5)$, $t(3)$ and $t(1)$:

$$\begin{cases} t(5)=-a+\sqrt[5]{a^5+b_5} \\ t(3)=-c+\sqrt[3]{c^3+b_3} \\ t(1)=\dfrac{B-A_5b_5-\left(A_3-A_5\times 10a^2\right)b_3}{A_1-A_5\times 5a^4-\left(A_3-A_5\times 10a^2\right)\times 3c^2} \end{cases}$$

In order to make the function (43) to be established, the value of $t(5)$, $t(3)$ and $t(1)$ must be equal, let $t(5)=t(3)=t(1)=m$, then,

$$m=\frac{B-A_5\left[(m+a)^5-a^5\right]-\left(A_3-A_5\times 10a^2\right)\left[(m+c)^3-c^3\right]}{A_1-A_5\times 5a^4-\left(A_3-A_5\times 10a^2\right)\times 3c^2} \quad (45)$$

The function (45) can be arranged as:

$$w^5+pw^3+qw^2+rw+s=0 \quad (46)$$

$$m+a=w$$

$$\begin{cases} p=\dfrac{A_3-A_5\times 10a^2}{A_5} \\ q=\dfrac{\left(A_3-A_5\times 10a^2\right)\times 3(c-a)}{A_5} \\ r=\dfrac{\left(A_3-A_5\times 10a^2\right)\times 3(c-a)^2+\left[A_1-A_5\times 5a^4-\left(A_3-A_5\times 10a^2\right)\times 3c^2\right]}{A_5} \\ s=\dfrac{\left(A_3-A_5\times 10a^2\right)(c-a)^3-a\left[A_1-A_5\times 5a^4-\left(A_3-A_5\times 10a^2\right)\times 3c^2\right]-B-A_5a^5-\left(A_3-A_5\times 10a^2\right)c^3}{A_5} \end{cases}$$

The function (46) can be deformed as:

$$w\left(w^4+pw^2+r\right)+\left(qw^2+s\right)=0 \quad (47)$$

Let $w^2=x$, then the function (47) can be deformed as:

$$\begin{aligned} w\left(x^2+px+r\right)+(qx+s)&=w(qx+s)(\alpha x+\beta)+(qx+s) \\ &=(qx+s)\left[w(\alpha x+\beta)+1\right] \\ &=\left(qw^2+s\right)\left(\alpha w^3+\beta w+1\right) \end{aligned} \quad (48)$$

If p, r, α, β satisfies the condition (49) below:

$$\begin{cases} q\alpha=1 \\ s\beta=r \\ s\alpha+q\beta=s\dfrac{1}{q}+q\dfrac{r}{s}=p \\ s^2+q^2r=pqs \end{cases} \quad (49)$$

Therefore, if p, q, r, s satisfy the function (49), the solution of the function (48) are easy to be found.

$$w_{1,2} = \pm\sqrt{-\frac{s}{q}}$$

$$w_{3,4,5} = w = \sqrt[3]{-\frac{q}{2}+\sqrt{\left(\frac{q}{2}\right)^2+\left(\frac{qr}{3s}\right)^3}}+\sqrt[3]{-\frac{q}{2}-\sqrt{\left(\frac{q}{2}\right)^2+\left(\frac{qr}{3s}\right)^3}}$$

But, if those conditions can't be satisfied, then the function (46) can't be deformed as function (47), so we can deform the function (46) to function (50):

$$\left(w^3+dw^2+ew+f\right)\left(w^2-dw+g\right)=0 \tag{50}$$

The unknown number d, e, f, g satisfy the relationship (51) below:

$$\begin{cases} e-d^2+g=p \\ f-de+dg=q \\ -df+eg=r \\ fg=s \end{cases} \tag{51}$$

There are four functions in (51), so we can solve each value of the unknown number d, e, f, g, then the solution for the random equation of degree 5 with one unknown (50) was found. However, to solve the function (51) is beyond my ability, therefore, I leave this puzzle to the clever reader.

For example, Let's find the solution of function (52):

$$t^5+2t^4+3t^3+4t^2+5t-6=0 \tag{52}$$

$$a=\frac{A_4}{5A_5}=\frac{2}{5}$$

$$c=\frac{A_2-A_5\times 10a^3}{3\left(A_3-A_5\times 10a^2\right)}=\frac{4-10\times\frac{8}{125}}{3\left(3-10\times\frac{4}{25}\right)}=\frac{4}{5}$$

$$p=\frac{A_3-A_5\times 10a^2}{A_5}=\frac{7}{5}$$

$$q=\frac{\left(A_3-A_5\times 10a^2\right)\times 3(c-a)}{A_5}=\frac{42}{25}$$

$$r=\frac{\left(A_3-A_5\times 10a^2\right)\times 3(c-a)^2+\left[A_1-A_5\times 5a^4-\left(A_3-A_5\times 10a^2\right)\times 3c^2\right]}{A_5}=\frac{357}{125}$$

$$s=\frac{\left(A_3-A_5\times 10a^2\right)(c-a)^3-a\left[A_1-A_5\times 5a^4-\left(A_3-A_5\times 10a^2\right)\times 3c^2\right]-B-A_5a^5-\left(A_3-A_5\times 10a^2\right)c^3}{A_5}=-7.49056$$

$$e-d^2+g=\frac{7}{5}$$

$$f-de+dg=\frac{42}{25}$$

$$-df+eg=\frac{357}{125}$$

$$fg=-7.49056 \tag{53}$$

If we can find the value of *d*, *e*, *f*, *g*, then the solution for the Equation (52) was found. Obviously, the value of *d*, *e*, *f*, *g* can't be positive integer, then the solution *t* of function (52) can't be positive integer.

Similarly, if $n=6$, the function (26) is equal to the function (54) shows below:

$$t^5+6at^4+15a^2t^3+20a^3t^2+15a^4t^1-\left(m-6a^5\right)=0 \tag{54}$$

$$a'=\frac{A_4}{5A_5}=\frac{6}{5}a$$

$$c'=\frac{A_2-A_5\times 10a'^3}{3\left(A_3-A_5\times 10a'^2\right)}=\frac{20a^3-10\times\frac{216}{125}a^3}{3\left(15a^2-10\times\frac{36}{25}a^2\right)}=-\frac{4}{15}a$$

$$p=\frac{A_3-A_5\times 10a'^2}{A_5}=\frac{3}{5}a^2$$

$$q=\frac{\left(A_3-A_5\times 10a'^2\right)\times 3\left(c'-a'\right)}{A_5}=-\frac{66}{25}a^3$$

$$r=\frac{\left(A_3-A_5\times 10a'^2\right)\times 3\left(c'-a'\right)^2+\left[A_1-A_5\times 5a'^4-\left(A_3-A_5\times 10a'^2\right)\times 3c'^2\right]}{A_5}=\frac{1047}{125}a^4$$

$$s=\frac{\left(A_3-A_5\times 10a'^2\right)\left(c'-a'\right)^3-a'\left[A_1-A_5\times 5a'^4-\left(A_3-A_5\times 10a'^2\right)\times 3c'^2\right]-B-A_5a'^5-\left(A_3-A_5\times 10a'^2\right)c'^3}{A_5}$$

$$=-\left(9.77472a^5+B\right)=-\left(3.77472a^5+m\right)$$

a and m are positive integer, if *p*, *q*, *r*, *s* are positive integer, then the solution *t* of function (54) can be a positive integer. Therefore, $a=5k$, k is positive integer, so:

$$p=\frac{3}{5}a^2=15k^2$$

$$q=-\frac{66}{25}a^3=-330k^3$$

$$r=\frac{1047}{125}a^4=5235k^4$$

$$s=-\left(3.77472a^5+m\right)=-\left(11796k^5+m\right)$$

$$e-d^2+g=15k^2$$

$$f-de+dg=-330k^3$$

$$-df+eg=5235k^4$$

$$fg=-\left(11796k^5+m\right) \tag{55}$$

If we can solve the function (51), then we can find the relationship between *d*, *e*, *f*, *g* and *k* through function (55), then we can according to the result in **Part one** to judge if the function (54) can have a positive integer solution.

However, the function (26) can be further transformed to the function (26-1)

$$t^{n-2}+C_n^{(n-1)}a^1t^{(n-3)}+\cdots+C_n^3a^{(n-3)}t^1+C_n^2a^{(n-2)}t^0+C_n^1a^{(n-1)}-\frac{m}{t}=0 \tag{26-1}$$

Thus, the function (54) can be deformed to function (56):

$$t^4+6at^3+15a^2t^2+20a^3t^1+15a^4-\frac{m-6a^5}{t}=0 \tag{56}$$

Let $\frac{m-6a^5}{t}=m'$, then we can use the result in **Part two** to prove the value of t in function (56) cannot be a positive integer. What's more, the function (56) can be transformed further to function (57):

$$t^3+6at^2+15a^2t^1+20a^3-\frac{m'-15a^4}{t}=0 \tag{57}$$

Let $\frac{m'-15a^4}{t}=m''$, then we can use the result in **Part one** to prove the value of t in function (57) cannot be a positive integer.

$$t^3+6at^2+15a^2t^1-\left(m''-20a^3\right)=0 \tag{57}$$

$$p=\frac{3A_1A_3-A_2^2}{3A_3^2}=3a^2$$

$$q=-\frac{27A_3^2B+9A_1A_2A_3-2A_2^3}{27A_3^3}=6a^3-m''$$

$$w=\sqrt[3]{-\frac{q}{2}+\sqrt{\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3}}+\sqrt[3]{-\frac{q}{2}-\sqrt{\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3}}$$

$$u=\sqrt{\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3}=\sqrt{\left(3a^3-\frac{m''}{2}\right)^2+\left(a^2\right)^3} \tag{58}$$

The value of u in the function (58) must be a integer, therefore, $\left(3a^3-\frac{m''}{2}\right)^2+\left(a^2\right)^3$ must be a positive square integer, otherwise, the value of w can't be a rational number.

Let $m''=2k$, k is positive integer, then:

$$u=\sqrt{\left(3a^3-\frac{m''}{2}\right)^2+\left(a^2\right)^3}=\sqrt{\left(3a^3-k\right)^2+\left(a^2\right)^3}$$

Obviously, if $\left(3a^3-k\right)^2=K^2a^6$, $K^2+1=R^2$, K and R is a positive integer, under this condition, the value of u can be a integer. Then, $K^2=0,3,8,15,\cdots$, $R=1,2,3,4,\cdots$, and $u=\pm Ra^3$. However, the value of K must be a positive square integer, a positive square integerplus 1 is still a square integer, only when $K^2=0,R^2=1$ can satisfy the requirement, then $k=3a^3$, $m''=2k=6a^3$,

$u = \pm a^3$.

$$w = \sqrt[3]{0 + \left(\pm a^3\right)} + \sqrt[3]{0 - \left(\pm a^3\right)} = 0$$

Therefore, the solution of *t* satisfy the function in (57) is $t = -\frac{A_2}{3A_3} + w = -2a + w = -2a$. However, the

$$\frac{m' - 15a^4}{t} = m'' = 6a^3$$

$$\frac{m - 6a^5}{t} = m'$$

$$\frac{b^n}{t} = m \tag{59}$$

According to the 3 equation infunction (59), we can get:

$$6a^3t^3 + 15a^4t^2 + 6a^5t^1 - b^n = 0 \tag{60}$$

Substitute $t = -2a$ into the function (60), then $b^n = 0$, thus $b = 0$, this is contradictory with $b > 0$, therefore, the value of *t* in function (57) can't be positive integer, thus, the Fermat's Last Theorem is established when $n = 6$.

Through this method, we can prove the Fermat's Last Theorem is established when $n = 6$ even we don't know the solution of equation of degree *5* with one unknown, what's more, we can use the same method to prove the Fermat's Last Theorem is established when $n > 6$, thus the Fermat's Last Theorem was proved.

Use the way we transform the function (26) to function (26-1), we can finally transform the function (26) to function (26-2):

$$t^3 + C_n^{(n-1)}a^1t^2 + C_n^{(n-2)}a^2t^1 - \left(m'' - C_n^{(n-3)}a^3\right) = 0 \tag{26-2}$$

$$m'' = \frac{m}{t^{n-4}} - \frac{C_n^1 a^{(n-1)}}{t^{n-4}} - \frac{C_n^2 a^{(n-2)}}{t^{n-5}} - \cdots - \frac{C_n^{(n-4)}a^4}{t^1}$$

The function (26-2) is equal to:

$$t^3 + na^1t^2 + \frac{n(n-1)}{2}a^2t^1 - \left(m'' - \frac{n(n-1)(n-2)}{6}a^3\right) = 0 \tag{26-2}$$

$$p = \frac{3A_1A_3 - A_2^2}{3A_3^2} = \frac{n(n-3)}{6}a^2$$

$$q = -\frac{27A_3^2B + 9A_1A_2A_3 - 2A_2^3}{27A_3^3} = \frac{n(n-3)(2n-3)}{27}a^3 - m''$$

$$w = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}$$

$$u = \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3} = \sqrt{\left(\frac{n(n-3)(2n-3)}{54}a^3 - \frac{m''}{2}\right)^2 + \left(\frac{n(n-3)}{18}a^2\right)^3} \tag{61}$$

The value of *u* in the function (61) must be a integer, therefore,

$\Delta=\left(\frac{n(n-3)(2n-3)}{54}a^3-\frac{m''}{2}\right)^2+\left(\frac{n(n-3)}{18}a^2\right)^3$ must be a positive square integer, otherwise, the value of w can't be a rational number.

Let $a=3s$, $m''=2ks^3$, s and k is positive integer, then:

$$u=\sqrt{\left(\frac{n(n-3)(2n-3)}{2}s^3-ks^3\right)^2+\left(\frac{n(n-3)}{2}s^2\right)^3}$$

$\frac{n(n-3)(2n-3)}{2}$ and $\frac{n(n-3)}{2}$ are integer when n is integer and $n>3$.

Let $\frac{n(n-3)}{2}=r$, and r is a positive integer ($r=2,5,9,14,20,\cdots$), then $\Delta=\left((2n-3)rs^3-ks^3\right)^2+\left(rs^2\right)^3$

Obviously, if $\left((2n-3)rs^3-ks^3\right)^2=K^2s^6$, $K^2+r^3=R^2$, K and R is integer, under this condition, the value of u can be a integer.

$$\frac{\Delta}{s^6}=\left((2n-3)r-k\right)^2+r^3=R^2=K^2+r^3$$

$$r^3=R^2-K^2=(R+K)(R-K)$$

Let $(R+K)=\frac{r^2}{m}$, then $(R-K)=mr$,

R, k and r are integer, then $m\in Q$

$$R=\left(\frac{r^2}{m}+mr\right)\Big/2$$

$$K=\left(\frac{r^2}{m}-mr\right)\Big/2$$

$$\frac{w}{s}=\sqrt[3]{-\frac{q}{2}+u}+\sqrt[3]{-\frac{q}{2}-u}=\sqrt[3]{-K+R}+\sqrt[3]{-K-R}=\sqrt[3]{mr}+\sqrt[3]{-\frac{r^2}{m}}$$

Only if $m=-\frac{N^3}{r}$, N is integer but $N\neq 0$, $r=\frac{n(n-3)}{2}$, then

$\frac{w}{s}=-N+\frac{r}{N}=-N+\frac{n(n-3)}{2N}$ is rational, then

$t=-\frac{n}{3}a+w=\left(-n-N+\frac{n(n-3)}{2N}\right)s$ can be integer.

$$R=\frac{\frac{r^2}{m}+mr}{2}=-\frac{\left(\frac{r}{N}\right)^3+N^3}{2}=-\frac{\left(\frac{n(n-3)}{2N}\right)^3+N^3}{2}$$

$$K=\frac{\frac{r^2}{m}-mr}{2}=-\frac{\left(\frac{r}{N}\right)^3-N^3}{2}=-\frac{\left(\frac{n(n-3)}{2N}\right)^3-N^3}{2}$$

The value of N must be the factor of r, thus the value of R, K could be integer.

For example, if n = 7, $r=\frac{n(n-3)}{2}=14$

$$R=-\frac{\left(\frac{14}{N}\right)^3+N^3}{2}$$

If N = 1, 2, 7, 14, $\left(\frac{14}{N}\right)^3+N^3$ is an odd number;

If N is integer and $N\neq 0$, 1, 2, 7, and 14, $\left(\frac{14}{N}\right)^3+N^3$ is a fraction number, therefore, $R=-\frac{\left(\frac{14}{N}\right)^3+N^3}{2}$ can not be a integer, which is contradictory with R is a integer. Thus $m\neq-\frac{N^3}{r}$ and $\frac{w}{s}$ is irrational, s is integer, the value of w is irrational, then the value of t in function (26-2) can't be positive integer when n = 7, so the Fermat's Last Theorem is established when n = 7.

Similarly, if n = 8, $r=\frac{n(n-3)}{2}=20$

$$R=-\frac{\left(\frac{20}{N}\right)^3+N^3}{2}$$

If N = 1, 2, 4, 5, 10, 20, $\left(\frac{20}{N}\right)^3+N^3$ can be a integer, but only N = 2, 10, R can be a integer.

When N = 2 or 10, R = −504, in other condition, R is not a integer.

$$K=\left((2n-3)r-k\right)=(260-k)=-\frac{\left(\frac{20}{N}\right)^3-N^3}{2}=\pm 496$$

$$r^3=R^2-K^2=20^3=(-504)^2-(\pm 496)^2$$

$$\frac{w}{s}=-N+\frac{r}{N}=\pm 8$$

$$t=-\frac{n}{3}a+w=\left(-n-N+\frac{n(n-3)}{2N}\right)s$$

When N = 2, $t=(-8-2+10)*s=0*s=0$

When N = 10, $t=(-8-10+2)*s=-16*s<0$

which is contradictory with t is a integer and $0<t<b$. so the Fermat's Last Theorem is established when n = 8.

If n = 9, $r=\frac{n(n-3)}{2}=27$

$$R=-\frac{\left(\frac{27}{N}\right)^3+N^3}{2}$$

If N = 1, 3, 9, 27, $\left(\frac{27}{N}\right)^3 + N^3$ can be even number, and R can be a integer.

N = 1, R = −9842, $K = -9841$, $\frac{w}{s} = 26$, $t = 17s$

N = 3, R = −378, $K = -351$, $\frac{w}{s} = 6$, $t = -3s$

N = 9, R = −378, $K = 351$, $\frac{w}{s} = 6$, $t = -15s$

N = 27, R = −9842, $K = 9841$, $\frac{w}{s} = -26$, $t = -35s$

Only if N = 1, $t = 17s$ satisfies the condition. What's more, we have to verify if the value of b is integer or not.

$$K = ((2n-3)r - k) = -9841, \quad k = 10246$$

$$m'' = 2ks^3 = 20592s^3$$

$$m'' = \frac{m}{t^{n-4}} - \frac{C_n^1 a^{(n-1)}}{t^{n-4}} - \frac{C_n^2 a^{(n-2)}}{t^{n-5}} - \cdots - \frac{C_n^{(n-4)} a^4}{t^1}$$

$\frac{b^n}{t} = m$, $a = 3s$, $n = 9$, $t = 17s$, then:

$$m'' = \frac{b^9}{(17s)^6} - \frac{9(3s)^8}{(17s)^5} - \frac{36(3s)^7}{(17s)^4} - \frac{84(3s)^6}{(17s)^3} - \frac{126(3s)^5}{(17s)^2} - \frac{126(3s)^4}{(17s)^1} = 20592s^3$$

$$\begin{aligned} b^9 &= 9(3s)^8 * (17s) + 36(3s)^7 * (17s)^2 + 84(3s)^6 * (17s)^3 \\ &\quad + 126(3s)^5 * (17s)^4 + 126(3s)^4 * (17s)^5 + 20592s^3 * (17s)^6 \\ &= 514413737217 * s^9 \end{aligned}$$

$$b = \sqrt[9]{514413737217}s$$

s is a positive integer, thus b is irrational, which is contradictory with b is a integer. Therefore, the Fermat's Last Theorem is established when n = 9.

If n = 10, $r = \frac{n(n-3)}{2} = 35$

Only if N = 1, $t = 24s$ satisfies the condition.

Use the same way to verify if b is a integer or not when N = 1, $t = 24s$,

$$\begin{aligned} b^{10} &= 10(3s)^9 (24s)^1 + 45(3s)^8 * (24s)^2 + 120(3s)^7 * (24s)^3 \\ &\quad + 210(3s)^6 * (24s)^4 + 252(3s)^5 * (24s)^5 \\ &\quad + 210(3s)^4 * (24s)^6 + 44064s^3 * (24s)^7 \\ &= 205891132035600 * s^{10} \end{aligned}$$

$$b = \sqrt[10]{205891132035600}s$$

s is a positive integer, thus b is irrational, which is contradictory with b is a integer. Therefore, the Fermat's Last Theorem is established when n = 10.

The values of b satisfy the condition under different n is concluded in the Table 1.

Table 1 indicated that if 2 < n < 21, in order to have a positive integer solution

Table 1. The value of b under different n.

n	r	N	k	t	b
3	0	none			
4	2	none			
5	5	1	97	$-s$	
6	9	1	445	$2s$	$\sqrt[6]{14896s}$
7	14	none			
8	20	2, 10	756, −236	0, $-16s$	
9	27	1	10,246	$17s$	$\sqrt[9]{514413737217s}$
10	35	1	22,032	$24s$	$\sqrt[10]{205891132035600s}$
11	44	2	6156	$9s$	$\sqrt[11]{743008193541s}$
12	54	none			
13	65	1	138,807	$51s$	$\sqrt[13]{33198531813531451984941s}$
14	77	1	230,191	$62s$	$\sqrt[14]{461540731532546208768008s}$
15	90	none			
16	104	2, 4	73,316, 11,772	$34s$, $6s$	$\sqrt[16]{12337511914217166319227520s}$ $\sqrt[16]{1853020145805120s}$
17	119	1	846,268	$101s$	$\sqrt[17]{19479004955562800041143429455772221s}$
18	135	1, 3, 5	1,234,642, 50,004, 14,234	$116s$, $24s$, $4s$	$\sqrt[18]{23015822943866122205667120134071432s}$ $\sqrt[18]{58149737003040059302969680s}$ $\sqrt[18]{1628413210489960s}$
19	152	2, 4	224,804, 32,724	$55s$, $15s$	$\sqrt[19]{3199866632452173458088314772999205s}$ $\sqrt[19]{708235345355336514096165s}$
20	170	none			

Note: "none" means there is no integer of N to make $R=-\frac{\left(\frac{r}{N}\right)^3+N^3}{2}$ and $K=-\frac{\left(\frac{r}{N}\right)^3-N^3}{2}$ to be integer.

of t in function (26-2), the value of b cannot be a integer (Although the value of b is very close to a integer), which means b and c($c = a + t$) in function (1) can not be integer at the same time, thus the Fermat's Last Theorem is established when $2 < n < 21$.

Use the same way, we can easily prove the Fermat's Last Theorem is established under different indices n.

4) Part four: The conclusion of Section 2.3

In order to prove the value of c satisfies the function (1) can't be positive integer when a, b, and n are positive integer and $n > 2$, in this section, we find the solution for the random equation of degree n with one unknown (when $n = 3$ and 4), and proved that the Fermat's theorem was established at $n = 3$, 4 and 5, the solution for the random equation of degree 5 with one unknown is hard to be found, if we find this solution, we can prove that the Fermat's theorem was established at n = 6. Therefore, if we want to prove the Fermat's theorem, we have to find the solution for the random equation of degree n with one unknown, then we can solve the certain function (26)

$$t^{n-1} + C_n^{(n-1)} a^1 t^{(n-2)} + \cdots + C_n^2 a^{(n-2)} t^1 + C_n^1 a^{(n-1)} - m = 0 \quad (26)$$

has no positive integer solution, thus the Fermat's theorem was proved.

However, if we can't find the solution for the random equation of degree n with one unknown when $n > 4$, we can transfer the function (26) into (26-2)

$$t^3 + C_n^{(n-1)} a^1 t^2 + C_n^{(n-2)} a^2 t^1 - \left(m'' - C_n^{(n-3)} a^3\right) = 0 \quad (26\text{-}2)$$

$$m'' = \frac{m}{t^{n-4}} - \frac{C_n^1 a^{(n-1)}}{t^{n-4}} - \frac{C_n^2 a^{(n-2)}}{t^{n-5}} - \cdots - \frac{C_n^{(n-4)} a^4}{t^1}$$

thus we proved the Fermat's Last Theorem is established when n = 3, 4, 5, ……, 20 in this section, and in this method, we can also prove the Fermat's Last Theorem is established when $n > 20$, then the Fermat's theorem was proved.

3. Extension of Fermat's Last Theorem

Based on the Fermat's Last Theorem, I put forward another extension theorem:

$$\sum_{i=1}^{i=k} a_i^n = a_1^n + a_2^n + \cdots + a_k^n = b^n \quad (\text{E-1})$$

$$\sum_{i=1}^{i=n} a_i^n = a_1^n + a_2^n + \cdots + a_n^n = b^n \quad (\text{E-2})$$

When n is integer and $n > 2$, to satisfy the function *(E*-1) and *(E*-2), $a_1, a_2, \cdots, a_k$ ($k = 2, 3, \cdots$, when k = 2, function *(E*-1) is equivalent to the Fermat's Last Theorem)and $a_1, a_2, \cdots, a_n$ ($n = 3, 4, \cdots$) and b can't be positive integer at the same time.

The function $a_1^3 + a_2^3 + a_3^3 = b^3$ has positive integer solution, so the key of the extension theorem is to find: under what conditions, the function *(E*-1) and *(E*-2) have no positive integer solution?

4. Conclusions

In this paper, I proposed an easy way to prove the Fermat's Last Theorem through a geometric method, and I found the relationship between $a^n + b^n = c^n = (a+t)^n$ and the solution t for equation of degree n with one unknown, then I found the solution when n = 3 and 4. If we can't find the solution of function (26) when $n > 5$,

$$t^{n-1} + C_n^{(n-1)} a^1 t^{(n-2)} + \cdots + C_n^2 a^{(n-2)} t^1 + C_n^1 a^{(n-1)} - m = 0 \quad (26)$$

$$m = \frac{b^n}{t}$$

We can transfer function (26) to (26-2)

$$t^3 + C_n^{(n-1)} a^1 t^2 + C_n^{(n-2)} a^2 t^1 - \left(m'' - C_n^{(n-3)} a^3\right) = 0 \tag{26-2}$$

$$m'' = \frac{m}{t^{n-4}} - \frac{C_n^1 a^{(n-1)}}{t^{n-4}} - \frac{C_n^2 a^{(n-2)}}{t^{n-5}} - \cdots - \frac{C_n^{(n-4)} a^4}{t^1}$$

In this way, we can easily prove the Fermat's Last Theorem is established under different indices n, and the value of b that make $c = a + t$ to be a integer under different n is:

When $n = 3$, $b^3 = 2ks^3 = 0$

When $n > 3$,

$$\begin{aligned} b^n &= Bs^n = (\zeta s)^n \\ &= C_n^{(n-1)} (3s)^{(n-1)} (t)^1 + C_n^{(n-2)} (3s)^{(n-2)} (t)^2 \\ &\quad + \cdots + C_n^4 (3s)^4 * (t)^{(n-4)} + 2ks^3 * (t)^{(n-3)} \end{aligned}$$

$$k = \frac{n(n-3)(2n-3)}{2} + \frac{\left(\frac{n(n-3)}{2N}\right)^3 + N^3}{2}$$

$$t = \left(-n - N + \frac{n(n-3)}{2N}\right)s$$

N is the factor of $r = \frac{n(n-3)}{2}$ that can make k and t to be a positive integer, s is a positive integer. $\zeta = \sqrt[n]{B}$ is irrational that very closing to a integer, thus $b = \zeta s$ cannot be a positive integer.

Therefore, the function $a^n + b^n = c^n$ can be established only under the following conditions:

1) When $\theta = 0$, $a + b = c$, $n = 1$

2) When $\theta = \frac{\pi}{2}$, $a^2 + b^2 = c^2$, $n = 2$

3) When $n > 2$, a, b, c are more than 0 and $a > b$: $\theta \in \left(0, \frac{\pi}{2}\right)$, $a^n + b^n < c^n$, $\theta \in \left(\frac{\pi}{2} + \arcsin\frac{b}{2a}, \pi\right)$, $a^n + b^n > c^n$, under this condition, the function $a^n + b^n = c^n$ cannot be established. $\theta \in \left(\frac{\pi}{2}, \frac{\pi}{2} + \arcsin\frac{b}{2a}\right)$, there is possible to make $a^n + b^n = c^n$, but the value of a, b and c can't be positive integer at the same time

Acknowledgements

I sincerely thank my family for trusting me and thank my friends for encouraging me. Especially, I thank the reader for reading this paper to the end and your questions are welcomed to be asked.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] https://www.daviddarling.info/encyclopedia/F/Fermats_last_theorem.html

[2] Taylor, R. and Wiles, A. (1995) Ring Theoretic Properties of Certain Hecke Algebras. *Annals of Mathematics*, **141**, 553-572. https://doi.org/10.2307/2118560 http://abel.math.harvard.edu/~rtaylor/

[3] Bratu, N.I. (2008) On the FERMAT'S Last Theorem. A New Proof for the Cases N-3 and N-5. *Octogan Math. Mag.*, **16**.

[4] https://www.math.net/article/fermats-last-theorem

[5] Wiles, A. (1995) Modular Elliptic-Curves and Fermat's Last Theorem. *Ann. Math.*, **141**, 443-551. https://doi.org/10.2307/2118559

[6] https://www.mathsisfun.com/pythagoras.html

[7] Euclid (2002) Euclid's Elements (Later Printing Edition). Green Lion Press, Santa Fe, NM.

[8] https://en.academic.ru/dic.nsf/enwiki/152736/