



Paper Type: Original Article

## Parthasarathian Transient Solution of M/M/1 Queue Manifold: Info-Geometric Analysis and Applications to Machine Learning

Ismail A Mageed\* 

AIMMA, IEEE, IAENG, School of Computer Science, AI, and Electronics, Faculty of engineering and Digital, Technologies, University of Bradford, United Kingdom; drismail664@gmail.com.

**Citation:**

<i>Received: 16 April 2024</i>	A Mageed, I. (2024). Parthasarathian transient solution of m/m/1 queue manifold: info-geometric analysis and applications to machine learning. <i>Optimality</i> , 1(2), 276-286.
<i>Revised: 13 June 2024</i>	
<i>Accepted: 18 August 2024</i>	

### Abstract


By constructing the Fisher Information Matrix (FIM) and its inverse (IFIM), the current study provides an info-geometric characterization of the transient M/M/1 queue manifold. Furthermore, stability's effect on IFIM's existence and its investigation of the Geodesic Equations (GEs) of motion have been made clear. Potentially even more, the article highlights several info-geometric machine learning applications. This in essence would manifest the huge contributions provided by this paper, which would in turn open new plethora of unlimited opportunities to the exploration of new trends in machine learning. In closing, some difficult open problems are discussed along with the next stage of the research.


**Keywords:** Transient M/M/queueuing system 1, Information geometry, Statistical manifold, Queue manifold, Information geodesic equations of motion, Fisher information matrix, Inverse fisher information matrix, Machine learning.

## 1 | Introduction

The Numerous study areas, including statistical sciences, have extensively used IG [1]. In other words, the goal of IG is to use statistics to apply the methods of Differential Geometry (DG), which indicates the major goal of IG is to use stochastic processes and probability theoretic in the applications of methodologies of non-Euclidean geometry.

IG supports SMs' descriptions that are based on intuitive reasoning. One might have a greater understanding of the crucial significance of IG [2]–[4].

 Corresponding Author: drismail664@gmail.com

 <https://doi.org/10.22105/opt.v1i2.60>



Licensee System Analytics. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).

Parametrization of a Statistical Manifold (SM) is visualized by Fig. 1 [4].

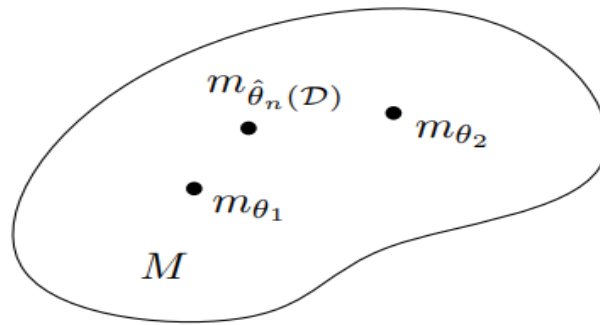


Fig. 1. SM's parametrization [3].

According to the literature, a paper by [5] examined info-geometrically the stable M/D/1 queue based on queue length routes 'characteristics, was the real motivator for the current study.

The main deliverables of this paper are described below.

- I. The discovery of the FIM and its inverse for the transient M/M/1 queue.
- II. Revealing the new discovery of the Geodesic Equations (GEs) of motion of the coordinates of the transient M/M/1 queue.
- III. A novel  $\alpha$ -connection [6] is introduced, which maps each coordinate to a value.
- IV. Highlighting potential IG applications to ML.

The remainder of the paper is divided into the following sections: preliminary IG definitions are given in Section 2. The transient M/M/1 QM's FIM and its IFIM are introduced in Section 3. While Section 4 gives the Information Geometric Equations of Motion (IMEs) for the coordinates of the transient M/M/1 QM. Section V addresses IG applications to ML. Section VI wraps up the paper and provides some challenging open problems combined with further research.

## 1.1 | Statistical Manifold

$M = \{p(x, \theta) | \theta \in \Theta\}$  is called an SM [6] if  $x$  is a random variable in sample space  $X$  and  $p(x, \theta)$  is the probability density function, which satisfies certain regular conditions.

## 2 | Potential Function

The potential function  $\Psi(\theta)$  [6] is the distinguished negative function of the coordinates alone of  $\mathcal{L}(x; \theta) = \ln(p(x; \theta))$ .

## 3 | Fisher's Information Matrix

Fisher's Information Matrix (FIM), namely  $[g_{ij}]$  reads as:

$$[g_{ij}] = \left[ \frac{\partial^2}{\partial \theta^i \partial \theta^j} (\Psi(\theta)) \right], i, j = 1, 2, \dots, n. \quad (1)$$

With respect to natural coordinates.

## 4 | Inverse Fisher Information Matrix (IFIM), Namely $[g^{ij}]$

$[g^{ij}]$  [6] reads:

$$[g^{ij}] = ([g_{ij}])^{-1} = \frac{\text{adj}[g_{ij}]}{\Delta}, \Delta = \det[g_{ij}]. \quad (2)$$

### 5 | $\alpha$ -Connection

For each real number  $\alpha$ , the  $\alpha$ - connection [7] reads:

$$\Gamma_{ij,k}^{(\alpha)} = \left(\frac{1-\alpha}{2}\right) (\partial_i \partial_j \partial_k(\Psi(\theta))). \tag{3}$$

With  $\Psi(\theta)$  Definition (2),  $\partial_i = \frac{\partial}{\partial \theta_i}$ .

### 6 | Geodesic Equations of Motion

The IEMs read as [6]

$$\frac{d^2 \theta^k}{dt^2} + \Gamma_{ij}^{k(0)} \left(\frac{d\theta^i}{dt}\right) \left(\frac{d\theta^j}{dt}\right) = 0, i, j = 1, 2, \dots, n, \quad \Gamma_{ij}^{k(\alpha)} = \Gamma_{ij,s}^{(\alpha)} g^{sk}. \tag{4}$$

By the above definition, the GEs are interpreted physically as the Information Geometric Equations of Motion, Shortly (IGEMs), or the Relativistic Equations of Motion (REMs), or the Riemannian equations of motion. At this stage, the current study provides a ground-breaking discovery of the IG analysis of transient queueing systems in comparison to that of non-time dependent queueing systems, namely IG analysis of stable queues [8], [9].

#### 6.1 | Fim And Ifim of the Transient M/M/1 Qm

The most straightforward probabilistic queueing model that can be studied analytically is a single-channel model with exponential inter-arrival durations, service times, and FIFO queue discipline. This is called the M/M/1 queue in queueing theory [10].

Based on the introduction of the function, Parthasarathy [10] suggested a straightforward method for the transient solution of the M/M/1 system to read as

$$p_n(t) = \rho p_0(t) e^{-(\lambda+\mu)t} \left[ \sum_{m=1}^n r_m(t) + e^{(\lambda+\mu)t} \right], \quad n = 0, 1, 2, \dots, \tag{5}$$

where  $\lambda$  stands for the mean rate per unit time at which arrival instants occur,  $\mu$  is the mean rate of service time,  $\rho = \frac{\lambda(t)}{\mu(t)}$  defines the traffic intensity or utilization factor.

$$r_n(t) = \left( \mu \beta_2^{n-a_1} (1 - \delta_{0a_1}) [I_{n-a_1}(\beta_1 t) - I_{n+a_1}(\beta_1 t)] + \lambda \beta_2^{n-1-a_1} (I_{n+1-a_1}(\beta_1 t) - I_{n-1-a_1}(\beta_1 t)) \right). \tag{6}$$

With  $\beta_1 = 2\sqrt{\lambda\mu}$ ,  $\beta_2 = 2\sqrt{\frac{\lambda}{\mu}}$ ,  $I_n(x)$  Eq. (19) is the modified Bessel function and

$$p_0(t) = \int_0^t r_1 e^{-(\lambda+\mu)y} dy + \delta_{0a_1}.$$

**Theorem 1.** The underlying queue of Eq. (5) has:

I. FIM reads as

$$[g_{ij}] = \begin{bmatrix} a & b & c \\ b & 0 & h \\ c & h & l \end{bmatrix}. \tag{7}$$

With

$$a = \frac{1}{\rho^2} - \frac{1}{(1-\rho)^2}, \quad (8)$$

$$b = t, \quad (9)$$

$$c = \frac{\rho}{\rho^2} + \frac{\rho}{(1-\rho)^2} + t\mu + \mu\rho, \quad (10)$$

$$h = (1 + \rho) + t\rho, \quad (11)$$

$$l = \left( \frac{\rho^2 - \rho\rho}{\rho^2} + \frac{\rho^2 + (1-\rho)\rho}{(1-\rho)^2} + ((1+\rho)t\mu + \mu + 2\rho\mu + 2\mu\rho + 2\rho\mu + \mu t\rho), \quad (12)$$

$$\Delta = \det([g_{ij}]) = (a(-h^2) - b(bl - ch) + c(bh)). \quad (13)$$

Provided that,  $\dot{\phantom{x}}$  refers to the temporal derivative  $\frac{d}{dt}$ .

II.  $[g^{ij}]$  reads as:

$$[g^{ij}] = \frac{\text{adj}[g_{ij}]}{\Delta} = \begin{bmatrix} A & B & L \\ B & E & F \\ L & F & H \end{bmatrix}, \quad (14)$$

where

$$A = \frac{(-h^2)}{\Delta}, \quad (15)$$

$$B = \frac{(ch - bl)}{\Delta}, \quad (16)$$

$$L = \frac{(bh)}{\Delta}, \quad (17)$$

$$E = \frac{(al - c^2)}{\Delta}, \quad (18)$$

$$F = \frac{(bc - ah)}{\Delta}, \quad (19)$$

$$H = \frac{(-b^2)}{\Delta}. \quad (20)$$

III. Following (5), we have

$$\mathcal{L}(x; \theta) = \ln(p_n(x; \theta)) = (\ln(\rho p_0(t)e^{-(\lambda+\mu)t}) + \ln(\rho p_0(t)e^{-(\lambda+\mu)t} [\sum_{m=1}^n r_m(t) + e^{(\lambda+\mu)t}])), \quad (21)$$

$$\theta = (\theta_1, \theta_2, \theta_3) = (\rho, \mu, t). \quad (22)$$

We have

$$\Psi(\theta) = -\ln(\rho) - \ln(1 - \rho) + \mu(1 + \rho)t, \quad p_0(t) = 1 - \rho. \quad (23)$$

Thus,

$$\partial_1 = \frac{\partial \Psi}{\partial \rho} = -\frac{1}{\rho} + \frac{1}{(1-\rho)} + \mu t, \quad \partial_2 = \frac{\partial \Psi}{\partial \mu} = (1 + \rho)t, \quad \partial_3 = \frac{\partial \Psi}{\partial t} = -\frac{\rho}{\rho} + \frac{\rho}{(1-\rho)} + (1 + \rho)t\mu + \mu(1 + \rho) + \mu t\rho, \quad (24)$$

$$\partial_1 \partial_1 = \frac{1}{\rho^2} - \frac{1}{(1-\rho)^2}, \quad (25)$$

$$\partial_1 \partial_2 = \partial_2 \partial_1 = t, \quad (26)$$

$$\partial_2 \partial_2 = 0, \quad (27)$$

$$\partial_1 \partial_3 = \frac{\rho}{\rho^2} + \frac{\rho}{(1-\rho)^2} + t\mu + \mu\rho = \partial_3 \partial_1, \quad (28)$$

$$\partial_2 \partial_3 = (1 + \rho) + t\rho = \partial_3 \partial_2, \quad (29)$$

$$\partial_3 \partial_3 = \left( \frac{\rho^2 - \rho\rho}{\rho^2} + \frac{\rho^2 + (1-\rho)\rho}{(1-\rho)^2} + (1+\rho)t\mu + \mu + 2\rho\mu + 2\mu\rho + 2\rho\mu t\rho \right). \quad (30)$$

Therefore, the fisher information matrix, FIM, is obtained (Eq. (7)).

$[g^{ij}]$  reads as:

$$\begin{aligned} [g^{ij}] &= [g_{ij}]^{-1} = \frac{\text{adj}[g_{ij}]}{\Delta} = \frac{\text{Transpose}(\text{Cov}[g_{ij}])}{\Delta} = \frac{1}{\Delta} \text{transpose}(\text{Cov} \begin{bmatrix} a & b & c \\ b & 0 & h \\ c & h & l \end{bmatrix}) \\ &= \begin{bmatrix} (-h^2) & (ch - bl) & (bh) \\ (ch - bl) & (al - c^2) & (bc - ah) \\ (bh) & (bc - ah) & (-b^2) \end{bmatrix}. \end{aligned} \quad (31)$$

Thus,

$$[g^{ij}] = \frac{1}{\Delta} \text{transpose}(\text{Cov} \begin{bmatrix} (-h^2) & (ch - bl) & (bh) \\ (ch - bl) & (al - c^2) & (bc - ah) \\ (bh) & (bc - ah) & (-b^2) \end{bmatrix}) = \begin{bmatrix} A & B & L \\ B & E & F \\ L & F & H \end{bmatrix}.$$

As requested, It is notable that FIM should satisfy the symmetry requirement.

In what follows, the components of  $\alpha$  (or  $\nabla^{(\alpha)}$ )- connection are obtained. These calculated expressions are needed to obtain the corresponding GEs of the parametric coordinates of M/M/1 QM.

From Eq. (3), the reader can check that.

$$\Gamma_{11,1}^{(\alpha)} = -(1 - \alpha) \left( \frac{1}{(1-\rho)^3} + \frac{1}{\rho^3} \right), \quad (32)$$

$$0 = \Gamma_{22,2}^{(\alpha)} = \Gamma_{12,2}^{(\alpha)} = \Gamma_{22,1}^{(\alpha)} = \Gamma_{21,2}^{(\alpha)} = \Gamma_{11,2}^{(\alpha)} = \Gamma_{12,1}^{(\alpha)} = \Gamma_{21,1}^{(\alpha)} = \Gamma_{22,3}^{(\alpha)} = \Gamma_{23,2}^{(\alpha)} = \Gamma_{32,2}^{(\alpha)}, \quad (33)$$

$$\Gamma_{31,1}^{(\alpha)} = (1 - \alpha) \rho \left( \frac{1}{(1-\rho)^3} - \frac{1}{\rho^3} \right), \quad (34)$$

$$\Gamma_{13,1}^{(\alpha)} = \Gamma_{11,3}^{(\alpha)} = (1 - \alpha) \rho \left( \frac{1}{(1-\rho)^3} - \frac{1}{\rho^3} \right), \quad (35)$$

$$\Gamma_{13,2}^{(\alpha)} = \Gamma_{12,3}^{(\alpha)} = \frac{(1 - \alpha)}{2}, \quad (36)$$

$$\Gamma_{13,3}^{(\alpha)} = \frac{(1 - \alpha)}{2} \left[ \frac{\rho}{\rho^2} + \frac{\rho}{(1-\rho)^2} + \frac{2\rho^2}{(1-\rho)^3} + t\mu + 2\mu \right], \quad (37)$$

$$\Gamma_{23,1}^{(\alpha)} = \Gamma_{21,3}^{(\alpha)} = \frac{(1 - \alpha)\rho}{2}, \quad (38)$$

$$\Gamma_{23,3}^{(\alpha)} = \frac{(1 - \alpha)(1 + 2\rho + t\rho)}{2}, \quad (39)$$

$$\Gamma_{31,2}^{(\alpha)} = \Gamma_{32,1}^{(\alpha)} = \frac{(1 - \alpha)}{2}. \quad (40)$$

Engaging the same procedure, the remaining components can be determined.

## 6.2 | The IGMES of the Underlying Queue Manifold

I. The IMEs of the server utilization coordinate,  $\rho$  of the transient M/M/1QM.

The IMEs corresponding to the arrival rate coordinate,  $\rho$  are

$$\frac{d^2\theta^1}{dt^2} + \Gamma_{ij}^{1(0)} \left( \frac{d\theta^i}{dt} \right) \left( \frac{d\theta^j}{dt} \right) = 0, i, j = 1, 2, 3.$$

Now, we are in a situation of trying to find the path of motion of family of families of IMEs corresponding to the server utilization coordinate,  $\rho$ .

$$\begin{aligned} & \left[ \frac{d^2\rho}{dt^2} + \left[ \frac{L}{2} + \frac{L\rho}{2} \right] \left( \frac{d\rho}{dt} \right) \left( \frac{d\mu}{dt} \right) + \left[ \frac{1}{2} \left( 2A\rho \left( \frac{1}{(1-\rho)^3} - \frac{1}{\rho^3} \right) + \right. \right. \\ & \left. \left. B + L \left[ \frac{\rho}{\rho^2} + \frac{\rho}{(1-\rho)^2} + \frac{2\rho^2}{(1-\rho)^3} + t\mu + 2\mu \right] \right) \right] + \\ & \frac{1}{2} \left( 2A\rho \left( \frac{1}{(1-\rho)^3} - \frac{1}{\rho^3} \right) + B + \left( \frac{\rho\rho - 2\rho^2}{\rho^3} + \frac{2\rho^2 + (1-\rho)\rho}{(1-\rho)^3} + \right. \right. \\ & \left. \left. t\mu + \rho\mu + \mu + \mu\rho \right) L \right) \left( \frac{d\rho}{dt} \right) \left( \frac{d\mu}{dt} \right) + \left[ \rho L \left( \frac{1}{(1-\rho)^3} - \frac{1}{\rho^3} \right) \right] \left( \frac{d\rho}{dt} \right)^2 + \\ & \left[ \frac{1}{2} (A\rho + B(1 + 2\rho + t\rho)) + \frac{1}{2} (A + (\rho + t\rho)L) \right] \left( \frac{d\mu}{dt} \right) \left( \frac{d\mu}{dt} \right) + \\ & \left[ \frac{1}{2} \left( \left( \frac{\rho\rho - 2\rho^2}{\rho^3} + \frac{2\rho^2 + (1-\rho)\rho}{(1-\rho)^3} + t\mu + \rho\mu + \mu + \mu\rho \right) A + (\rho + \right. \right. \\ & \left. \left. t\rho) B + \left[ \frac{(\rho\rho - \rho\rho)\rho - 2\rho(\rho^2 - \rho\rho)}{\rho^3} + \frac{(\rho\rho + (1-\rho)\rho)(1-\rho) + 2\rho(\rho^2 + (1-\rho)\rho)}{(1-\rho)^3} + \right. \right. \right. \\ & \left. \left. (1 + \rho)t\mu + (1 + \rho)\mu + 3\rho\mu + \mu + 2\rho\mu + 2\mu\rho + \right. \right. \\ & \left. \left. 4\mu\rho + 2\rho\mu + \mu t\rho + \mu\rho + \mu t\rho \right] L \right) \left( \frac{d\mu}{dt} \right)^2 \right] = 0. \end{aligned} \quad (41)$$

It can be verified that one of the closed form solutions of Eq. (41) is determined by the paths of motion:

$$\rho(t) = \text{constant}, \mu(t) = \text{constant}, \text{ such that } \rho \in (0, 1), \mu > 1. \quad (42)$$

II. The IMEs of the Mean service Rate coordinate,  $\mu$  of the transient M/M/1 QM.

The IMEs of  $\mu$  are

$$\frac{d^2\theta^2}{dt^2} + \Gamma_{ij}^{1(0)} \left( \frac{d\theta^i}{dt} \right) \left( \frac{d\theta^j}{dt} \right) = 0, i, j = 1, 2, 3.$$

Now, we are in a situation of trying to find the path of motion of family of families of IMEs corresponding to the mean service rate,  $\mu$ .

$$\begin{aligned} & \left[ \frac{d^2\mu}{dt^2} + \left[ \frac{F}{2} + \frac{F\rho}{2} \right] \left( \frac{d\rho}{dt} \right) \left( \frac{d\mu}{dt} \right) + \left[ \frac{1}{2} \left( 2B\rho \left( \frac{1}{(1-\rho)^3} - \frac{1}{\rho^3} \right) + E + F \left[ \frac{\rho}{\rho^2} + \frac{\rho}{(1-\rho)^2} + \frac{2\rho^2}{(1-\rho)^3} + t\mu + \right. \right. \right. \\ & \left. \left. 2\mu \right] \right) + \frac{1}{2} \left( 2B\rho \left( \frac{1}{(1-\rho)^3} - \frac{1}{\rho^3} \right) + E + \left( \frac{\rho\rho - 2\rho^2}{\rho^3} + \frac{2\rho^2 + (1-\rho)\rho}{(1-\rho)^3} + t\mu + \rho\mu + \mu + \mu\rho \right) \right. \\ & \left. F \right) \left( \frac{d\mu}{dt} \right) \left( \frac{d\mu}{dt} \right) + \left[ \rho \left( \frac{1}{(1-\rho)^3} - \frac{1}{\rho^3} \right) F \right] \left( \frac{d\rho}{dt} \right)^2 + \left[ \frac{1}{2} (B\rho + E(1 + 2\rho + t\rho)) + \frac{1}{2} (B + (\rho + \right. \right. \\ & \left. \left. t\rho) F) \right] \left( \frac{d\rho}{dt} \right) \left( \frac{d\mu}{dt} \right) + \left[ \frac{1}{2} \left( \left( \frac{\rho\rho - 2\rho^2}{\rho^3} + \frac{2\rho^2 + (1-\rho)\rho}{(1-\rho)^3} + t\mu + \rho\mu + \mu + \mu\rho \right) B + (\rho + \right. \right. \\ & \left. \left. t\rho) E + \left[ \frac{(\rho\rho - \rho\rho)\rho - 2\rho(\rho^2 - \rho\rho)}{\rho^3} + \frac{(\rho\rho + (1-\rho)\rho)(1-\rho) + 2\rho(\rho^2 + (1-\rho)\rho)}{(1-\rho)^3} + \right. \right. \right. \end{aligned}$$

Thus,

$$\begin{aligned}
 & \left[ \frac{d^2 \mu}{dt^2} + \frac{1}{2} \left( \frac{\left( \frac{1}{\rho^2} - \frac{1}{(1-\rho)^2} \right) [((1+\rho)t\mu + \mu + 2\rho\mu) - (t\mu)^2]}{\left( \frac{1}{\rho^2} - \frac{1}{(1-\rho)^2} \right) (-((1+\rho)^2) - (t^2 [((1+\rho)t\mu + \mu + 2\rho\mu])]} \right) + \right. \\
 & \left. \frac{(t(t\mu) - \left( \frac{1}{\rho^2} - \frac{1}{(1-\rho)^2} \right) (1+\rho))}{\left( \frac{1}{\rho^2} - \frac{1}{(1-\rho)^2} \right) (-((1+\rho)^2) - (t^2 [((1+\rho)t\mu + \mu + 2\rho\mu])]} [t\mu + 2\mu] \right) + \right. \\
 & \left. \frac{(t(t\mu) - \left( \frac{1}{\rho^2} - \frac{1}{(1-\rho)^2} \right) (1+\rho))}{\left( \frac{1}{\rho^2} - \frac{1}{(1-\rho)^2} \right) (-((1+\rho)^2) - (t^2 [((1+\rho)t\mu + \mu + 2\rho\mu])]} \right) + \right. \\
 & \left. (t\mu + \mu) \frac{(t(t\mu) - \left( \frac{1}{\rho^2} - \frac{1}{(1-\rho)^2} \right) (1+\rho))}{\left( \frac{1}{\rho^2} - \frac{1}{(1-\rho)^2} \right) (-((1+\rho)^2) - (t^2 [((1+\rho)t\mu + \mu + 2\rho\mu])]} \right) \left] \left( \frac{d\mu}{dt} \right) + \frac{1}{2} ((t\mu + \mu) \right. \\
 & \left. \mu) \frac{(t\mu(1-\rho) - t[((1+\rho)t\mu + \mu])}{\left( \frac{1}{\rho^2} - \frac{1}{(1-\rho)^2} \right) (-((1+\rho)^2) - (t^2 [((1+\rho)t\mu + \mu + 2\rho\mu])]} + \right) ] = 0.
 \end{aligned} \tag{43}$$

It is obvious that for arbitrary constant values of  $\rho, \mu$ , we have a closed form solution for Eq. (43). As time becomes sufficiently large, i.e.,  $t \rightarrow \infty$ , Eq. (42) reduces to

$$\left[ \frac{3\mu}{2} - \frac{3\mu^2}{2(1+\rho)} - \frac{\mu\mu}{\mu} \right] = 0. \tag{44}$$

We propose the closed form solution,  $\mu = \frac{2\gamma t}{(1+\gamma t)}$ , for some arbitrary non-zero constant  $\gamma$ . substituting in Eq. (44) implies:

$$18\gamma^3 + \frac{6\gamma^2}{(1+\gamma t)(1+\rho)} + 6\gamma^4 = 0. \tag{45}$$

As  $t \rightarrow \infty$ ,  $\gamma = 0, 0, -3$ . The only accepted value is  $\gamma = -3$ ,  $\mu = \frac{6t}{(3t-1)}$ , which tends to the value  $\mu = 2$  as  $t \rightarrow \infty$ .

### III. The IMES of the temporal coordinate, $t$ for the underlying QM.

Engaging a similar lengthy mathematical manipulation, it can be verified that the IMES of the temporal coordinate,  $t$ , at infinite time is characterized by the family of families of temporal curves:

$$\mu(t) = \zeta_1 + \zeta_2 t + \zeta_3 t^2. \tag{46}$$

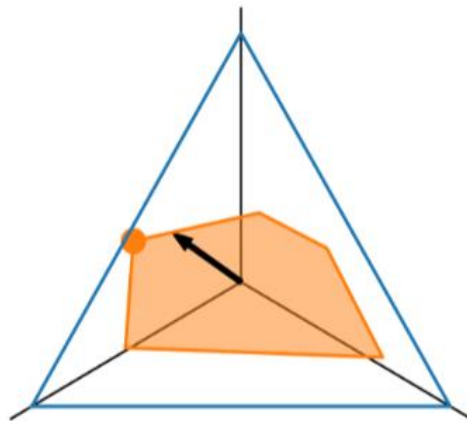
for some arbitrary non-zero constants  $\zeta_1, \zeta_2, \zeta_3, \zeta_1 > 1$ .

### IV. Ig Applications To ML

Stochastic gradient descent is a popular optimisation approach in ML that seeks to determine which direction of descent is the steepest within probability distributions' parameter space. Amari et al. [11] created the natural gradient, which is superior to the ordinary gradient since it more accurately captures this direction. Furthermore, even with gradient descent, FIM can be used to calculate the predicted change in output about parameter changes. Artificial Neural Networks (ANNs) are functions with multiple parameters that map input vectors to output vectors. They are widely applied in regression, computer vision, and speech processing. ANNs can be trained through supervised learning, where a training set of input-output pairs is used, or through unsupervised learning, which does not require a training set. The back-propagation algorithm efficiently computes the gradient descent by determining the contribution of each parameter to the error. Overfitting, a common issue in ANNs, can be mitigated by techniques such as using a validation set, early stopping, data augmentation, and dropout [12].

In Reinforcement Learning (RL) [13], when an agent needs to prepare for unknown tasks, one approach is unsupervised skill discovery. These algorithms learn a set of policies without using a reward function and are like representation learning algorithms in supervised learning. While prior work has shown that these methods can accelerate downstream tasks, our analysis reveals that they do not learn skills that are optimal for every possible reward function. However, the distribution over skills can provide an optimal initialization that minimizes regret against adversarially chosen reward functions, assuming a specific adaptation procedure. Notably, understanding which state marginals are achievable is crucial for analyzing the behavior of unsupervised skill learning algorithms [13] and their relevance to state-dependent reward functions. The set of achievable state marginal distributions can be characterized by a set of linear equations or by a complementary description based on a specific property.

This alternative perspective provides insights into the relationship between learned skills and downstream tasks in RL. So, if we have a set of possible state distributions, any combination of those distributions within the convex hull is also possible. This means that there exists a policy that can achieve any state distribution within that convex hull. By taking the convex hull of the state distributions of all deterministic policies, we can obtain the set of all possible state distributions, which is referred to as a polytope. Each vertex of the polytope represents a deterministic policy, and any policy's state distribution can be expressed as a combination of the state distributions at the vertices, as visualized by *Fig. 2*.



**Fig. 2. Representation of reward functions as vectors from the origin.**

The objective of maximizing the expected return is equivalent to maximizing the dot product between the state marginal distribution and the reward vector. This visualization helps us understand how different reward-maximizing policies are related to the state marginal polytope and provides insights into the relationship between state-dependent reward functions and optimal policies [13].

The analysis of where the skills learned by Mutual Information-based Skill Learning (MISL) are positioned on the state marginal polytope, which has motivated [13] to undertake this analysis by examining the mutual information objective and dissecting it to understand how the learned skills align with the polytope. In the context of skill learning [13], maximizing mutual information is equivalent to minimizing the maximum difference between a prior distribution over states (represented by a green square) and any possible state distribution.

This concept, known as information geometry, helps optimize skill-learning algorithms by assigning higher probabilities to policies with state distributions that deviate further from the average state distribution. The goal is to find skills as distinct as possible from each other and the average state distribution. Most importantly, the opaque circles represent the skills discovered by the Mutual Information Skill Learning (MISL) algorithm.

The dashed line represents equidistant state marginals from the green square, indicating an equal divergence from the prior distribution over states. This visualization helps illustrate the relationship between the skills learned by MISL and their distribution in relation to the average state distribution. This is illustrated in *Fig. 3*.



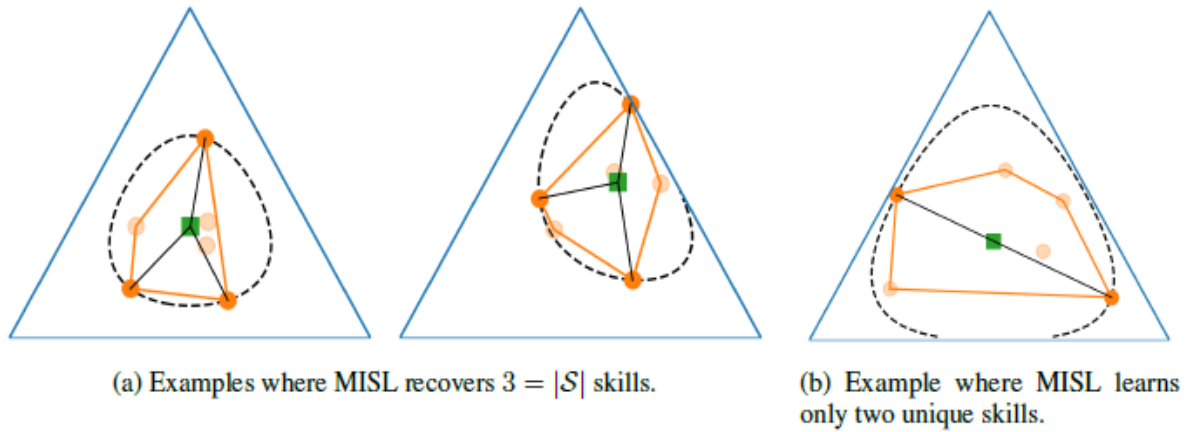


Fig. 3. Skill Learning's IG [13].

Reliable detection of Out-Of-Distribution (OOD) samples [14] is crucial for ensuring the safety of modern machine learning systems. The work of [14] introduced IGEOOD, an effective method that can detect OOD samples using any pre-trained neural network, regardless of the level of access to the ML model. By leveraging the geodesic distance between data distributions, IGEOOD combines confidence scores from the neural network's logits outputs and learned features to achieve superior performance compared to other state-of-the-art methods across different network architectures and datasets [14].

The authors [14] introduced the concept of IG and derived a metric based on the Fisher-Rao distance to measure the mismatch between probability density functions of a pre-trained neural classifier. Experimental results demonstrate that IGEOOD performs competitively with state-of-the-art methods in different setups, including black-box and grey-box scenarios, and achieves improved performance on various benchmarks. Fig. 4 showcases a comparison between Fisher-Rao and Mahalanobis distances for distinguishing between 1D Gaussian distributions in the context of OOD detection. It highlights the motivation to use the Fisher-Rao metric, which measures the dissimilarity between probability models, as it shows better performance in distinguishing between in-distribution and OOD samples compared to other distance-based approaches.

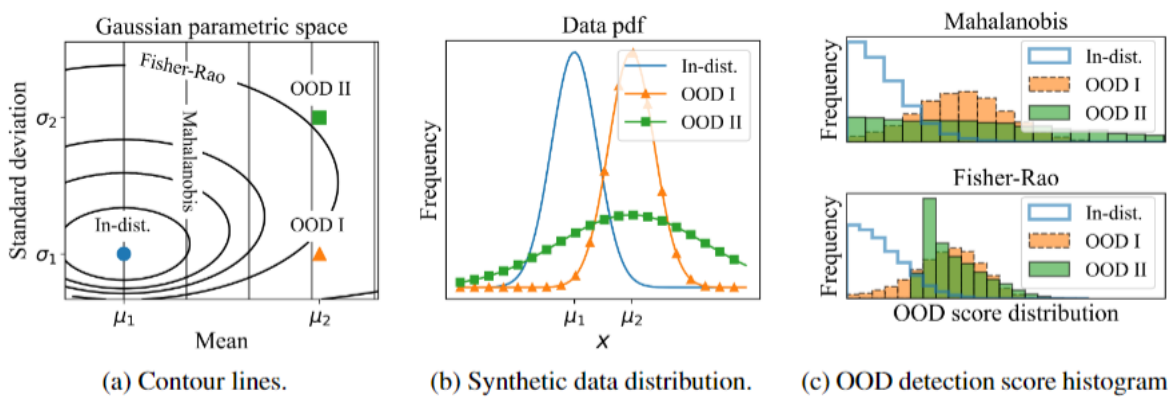


Fig. 4. Portraying the reason behind using Fisher-Rao OOD detection [14].

In the context of the experiment, the authors [14] compared three different settings for OOD detection. They use a unified metric, but with different formulations based on the type of distribution. For the Deep Neural Network (DNN) outputs, they use the softmax posterior probability distribution, while for the intermediate layers, they employ a model based on diagonal Gaussian Probability Density Functions (PDFs). This unified framework combines a single distance measure for both the softmax outputs and the latent features of the neural network, contributing to the separation of in-distribution and OOD samples, as illustrated in Fig. 5.

The distributions are shown for three different settings, specifically for a pre-trained DenseNet model on CIFAR-10 dataset, considering both in-distribution and OOD data from TinyImageNet (downsampled). This analysis helps evaluate the effectiveness of the IGEOOD score in distinguishing between in-distribution and OOD samples [13].

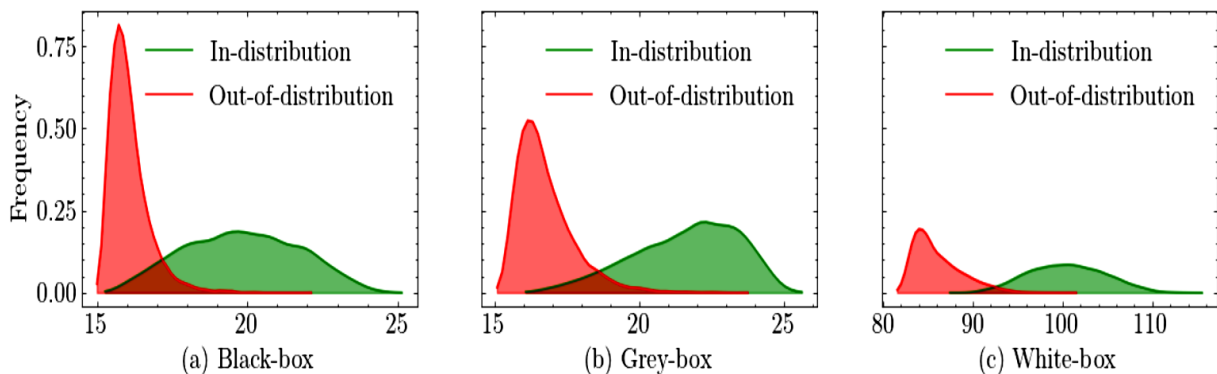


Fig. 5. The probability distributions of the IGEOOD score, which is a metric used for OOD detection.

## 7 | Conclusion

This study offers a revolutionary info-geometrics of transient M/M/1 QM. For this queue, FIM and IFIM are established. The GEs of motion for the queue's coordinates are determined. More potentially, some IG applications to ML are provided. Here are some challenging open problems to be addressed.

Open problem one: following the analysis in [13], a simplified model of skill learning algorithms that treated policy parameters and latent codes as a single representation, with different skills having distinct parameter sets was employed. However, practical implementations of skill learning algorithms are less flexible, which can alter the shape of the state marginal polytope.

This limited flexibility may lead to incomplete representation of certain state distributions, resulting in gaps or divisions in the polytope. It is yet to be determined if these practical implementations learn fewer unique skills and how these limitations impact the average state distribution, which is a really challenging open problem.

Open problem two: following the derivation of the potential function, PF.

$$\Psi(\theta) = -\ln(\rho) - \ln(1 - \rho) + \mu(1 + \rho)t, \quad p_0(t) = 1 - \rho.$$

Is the process of finding the threshold of PF based on its parameters decidable? The problem is still open.

Open problem three: having discovered PF (Eq. (21)), is it feasible to examine its algebraic properties as well as doing the same for its inverse, if it exists? This is a challenging open problem, yet unsolved so far.

Open problem four: the discovery of Revolutionary relativistic connections with the investigated transient queue is still unsolvable problem, for example finding the corresponding Gaussian, Ricci, Scalar and Einsteinian tensors for the transient M/M/1 QM.

The frontiers are open for unlimited explorations. The next phase of research includes answering the above open research problems and exploring more new avenues of IG applications to other scientific disciplines. More importantly, the possibility of employing Riemannian Geometric (RG) analysis, and the Theory of Relativity (TR) to the analysis of the dynamics of transient queues.

## Conflict of Interest

The authors declare no conflict of interest.

## Data Availability

All data are included in the text.

## Funding

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## References

- [1] Mageed, I. A., Zhou, Y., Liu, Y., & Zhang, Q. (2023). Towards a revolutionary info-geometric control theory with potential applications of fokker planck kolmogorov (fpk) equation to system control, modelling and simulation [presentation]. 2023 28th international conference on automation and computing (icac) (pp. 1–6). <https://doi.org/10.1109/ICAC57885.2023.10275271>
- [2] Skoda, Z. (2019). *Information geometry*. <https://ncatlab.org/nlab/show/information+geometry>
- [3] Mageed, I. A., Zhang, Q., Akinci, T. C., Yilmaz, M., & Sidhu, M. S. (2022). Towards abel prize: the generalized brownian motion manifold's fisher information matrix with info-geometric applications to energy works. *2022 global energy conference (GEC)* (pp. 379–384). IEEE. DOI: 10.1109/GEC55014.2022.9987239
- [4] Mageed, I. A., Yin, X., Liu, Y., & Zhang, Q. (2023). Za,b of the stable five-dimensional M/G/1 queue manifold formalism's info- geometric structure with potential info-geometric applications to human computer collaborations and digital twins. *28th international conference on automation and computing (ICAC)* (pp. 1–6). IEEE. DOI: 10.1109/ICAC57885.2023.10275185
- [5] Nakagawa, K. (2002). The geometry of M/D/1 queues and Large Deviation. *International transactions in operational research*, 9(2), 213–222.
- [6] Peng, L., Sun, H., & Jiu, L. (2007). The geometric structure of the Pareto distribution. *Boletín de la asociación matemática venezolana*, 14(1–2), 5–13.
- [7] Dodson, C. T. J. (2005). *Topics in information geometry*. The University of Manchester.
- [8] Mageed, I. A., & Kouvatsos, D. D. (2019). Information geometric structure of stable M/G/1 queue manifold and its matrix exponential. *35th uk performance engineering workshop 16 december 2019* (p. 116). IEEE. [https://www.researchgate.net/profile/Dr-Vijay-Singh-2/publication/340951179\\_proceedings/links/5ea70fb145851553fab34b07/proceedings.pdf#page=123](https://www.researchgate.net/profile/Dr-Vijay-Singh-2/publication/340951179_proceedings/links/5ea70fb145851553fab34b07/proceedings.pdf#page=123)
- [9] Mageed, I. A., & Kouvatsos, D. D. (2021). *The impact of information geometry on the analysis of the stable M/G/1 queue manifold*. ICORES (pp. 153–160). DOI: 10.5220/0010206801530160
- [10] Parthasarathy, P. R. (1987). A transient solution to an M/M/1 queue: a simple approach. *Advances in applied probability*, 19(4), 997–998.
- [11] Amari, S., Ozeki, T., Karakida, R., Yoshida, Y., & Okada, M. (2017). Dynamics of learning in MLP: Natural gradient and singularity revisited. *Neural computation*, 30(1), 1–33. <https://bsi-ni.brain.riken.jp/database/file/349/349.pdf>
- [12] Mageed, I. A., & Zhang, K. Q. (2022). Information geometry? exercises de styles. *Electronic journal of computer science and information technology*, 8(1), 9–14.
- [13] Eysenbach, B., Salakhutdinov, R., & Levine, S. (2021). *The information geometry of unsupervised reinforcement learning*. <https://arxiv.org/abs/2110.02719>
- [14] Gomes, E. D. C., Alberge, F., Duhamel, P., & Piantanida, P. (2022). *Igeood: an information geometry approach to out-of-distribution detection*. <https://arxiv.org/abs/2203.07798>