
Info-Geometric Analysis of the Dynamics of Parthasarathian Transient Solution of M/M/1 Queue Manifold with Info-Geometric Applications to Machine Learning

[Ismail A Mageed](#)*

Posted Date: 29 January 2024

doi: 10.20944/preprints202401.2008.v1

Keywords: Transient M/M/1; Information Geometry (IG); statistical manifold (SM); queue manifold (QM); Information Geodesic equations of motion (IEMs); Fisher Information matrix (FIM); Inverse Fisher Information matrix (IFIM); Machine Learning (ML)



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Info-Geometric Analysis of the Dynamics of Parthasarathian Transient Solution of M/M/1 Queue Manifold with Info-Geometric Applications to Machine Learning

Ismail A Mageed

School of Computer Science, AI, and Electronics, University of Bradford, United Kingdom;
drismailamageed@yahoo.com

Abstract: The current study characterizes the transient M/M/1 queue manifold info-geometrically, through devising Fisher Information matrix(FIM) and its inverse(IFIM). Additionally, the impact of stability on the existence of IFIM and explore the geodesic equations of motion has been revealed. More potentially, the paper discusses some info-geometric applications to Machine Learning are highlighted. Closing remarks with some challenging open problems combined with the next phase of research are given.

Keywords: Transient M/M/1 queueing system; Information Geometry (IG); statistical manifold (SM); queue manifold(QM); Information Geodesic equations of motion(IEMs); Fisher Information matrix (FIM); Inverse Fisher Information matrix (IFIM); Machine Learning (ML)

1. Introduction

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.

A manifold is a topological finite-dimensional Cartesian space, \mathbb{R}^n , in which an infinite-dimensional manifold exists[2]. Additionally, IG supports SMs' descriptions that are based on intuitive reasoning. One might have a greater understanding of the crucial significance of IG[3,4].

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

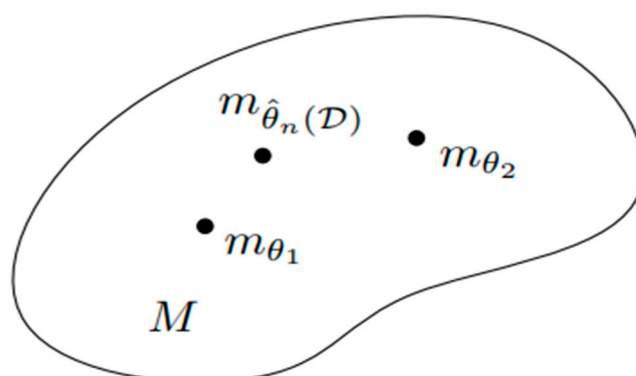


Figure 1. SM's parametrization [3].

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.

- The discovery of the FIM and its inverse for the transient M/M/1 queue.
- Revealing the new discovery of the geodesic equations of motion of the coordinates of the transient M/M/1 queue.
- In the context of the research paper, a novel α -connection [6] is introduced, which maps each coordinate to a value.
- Highlighting potential IG applications to ML.

The remainder of the paper is divided into the following sections: Preliminary IG definitions are given in Section II. The transient M/M/1 QM's FIM and its IFIM are introduced in Section III. While Section IV 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.

2. Main IG Definitions

1 Statistical Manifold (SM)

$M = \{p(x, \theta) | \theta \in \Theta\}$ is called an SM [7] if x is a random variable in sample space X and $p(x, \theta)$ is the probability density function, which satisfies certain regular conditions. Here, $\theta = (\theta_1, \theta_2, \dots, \theta_n) \in \Theta$ is an n -dimensional vector in some open subset $\Theta \subset \mathbb{R}^n$, and θ can be viewed as the coordinates on manifold M .

2 Potential Function

The potential function $\Psi(\theta)$ [7] is the distinguished negative function of the coordinates alone of $\mathcal{L}(x; \theta) = \ln(p(x; \theta))$

3 Fisher's Information Matrix (FIM)

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 IFIM

Given the FIM, the inverse matrix of $[g_{ij}]$ is defined by [7]

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

5 α -Connection

For each $\alpha \in \mathbb{R}$, the α (or $\nabla^{(\alpha)}$)-connection [6] is the torsion-free affine connection with components:

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

where $\Psi(\theta)$ is the potential function and $\partial_i = \frac{\partial}{\partial \theta^i}$.

6 Geodesic equations of motion

(1) The geodesic equations of manifold M with coordinate system $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ read as [7]

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

By the above definition, the geodesic equations 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].

3. Fim And IFIM of the Transient M/M/1QM

A single-channel model with exponential inter-arrival times, service times, and FIFO queue discipline is the simplest probabilistic queueing model that can be addressed analytically. In queueing theory, this is referred to as the M/M/1 queue [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

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

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

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

with $\beta_1 = 2\sqrt{\lambda\mu}$, $\beta_2 = 2\sqrt{\frac{\lambda}{\mu}}$, $I_n(x)$ [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 (5) has:

(i) FIM reads as

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

where

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

$$b = t (7)$$

$$c = \frac{\rho}{\rho^2} + \frac{\rho}{(1-\rho)^2} + t\mu + \mu\rho (8)$$

$$h = (1 + \rho) + t\rho (9)$$

$$l = \left(\frac{\rho^2 - \rho\rho}{\rho^2} + \frac{\rho^2 + (1-\rho)\rho}{(1-\rho)^2} + \right.$$

$$\left. ((1 + \rho)t\mu + \mu + 2\rho\mu + 2\mu\rho + 2\rho\mu + \mu t\rho) \right) (10)$$

$$\Delta = \det([g_{ij}]) = (a(-h^2) - b(bl - ch) + c(bh)) (11)$$

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

(ii) $[g^{ij}]$ reads as

$$[g^{ij}] = \frac{adj[g_{ij}]}{\Delta} = \begin{bmatrix} A & B & L \\ B & E & F \\ L & F & H \end{bmatrix} (12)$$

where

$$A = \frac{(-h^2)}{\Delta} (13)$$

$$B = \frac{(ch-bl)}{\Delta} (14)$$

$$L = \frac{(bh)}{\Delta} \quad (15)$$

$$E = \frac{(al-c^2)}{\Delta} \quad (16)$$

$$F = \frac{(bc-ah)}{\Delta} \quad (17)$$

$$H = \frac{(-b^2)}{\Delta} \quad (18)$$

Proof. (i) Following (5), we have:

$$\mathcal{L}(x; \theta) = \ln(p_n(x; \theta)) = (\ln(\rho p_0(t)e^{-(\lambda+\mu)t}) + \ln(\rho p_0(t)e^{-(\lambda+\mu)t} [\sum_{m=1}^n r_m(t) + e^{(\lambda+\mu)t}])) \quad (19)$$

$$\theta = (\theta_1, \theta_2, \theta_3) = (\rho, \mu, t) \quad (20)$$

We have

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

$$p_0(t) = 1 - \rho \quad (21)$$

Thus,

$$\partial_1 = \frac{\partial \Psi}{\partial \rho} = -\frac{1}{\rho} + \frac{1}{(1-\rho)} + \mu t, \partial_2 = \frac{\partial \Psi}{\partial \mu} = (1+\rho)t, \partial_3 = \frac{\partial \Psi}{\partial t} = -\frac{\rho}{\rho} + \frac{\rho}{(1-\rho)} + (1+\rho)t\mu + \mu(1+\rho) + \mu t\rho \quad (22)$$

$$\partial_1 \partial_1 = \frac{1}{\rho^2} - \frac{1}{(1-\rho)^2} \quad (23)$$

$$\partial_1 \partial_2 = \partial_2 \partial_1 = t \quad (24)$$

$$\partial_2 \partial_2 = 0 \quad (25)$$

$$\partial_1 \partial_3 = \frac{\rho}{\rho^2} + \frac{\rho}{(1-\rho)^2} + t\mu + \mu\rho = \partial_3 \partial_1 \quad (26)$$

$$\partial_2 \partial_3 = (1+\rho) + t\rho = \partial_3 \partial_2 \quad (27)$$

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

$$\left. (1+\rho)t\mu + \mu + 2\rho\mu + 2\mu\rho + 2\rho\mu + \mu t\rho \right) \quad (28)$$

Therefore, the Fisher Information Matrix, FIM, is obtained (c.f., (7)).

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

$$[g^{ij}] = [g_{ij}]^{-1} = \frac{adj[g_{ij}]}{\Delta} = \frac{Transpose(Cov[g_{ij}])}{\Delta} =$$

$$\frac{1}{\Delta} \text{transpose}(Cov \begin{bmatrix} a & b & c \\ b & 0 & h \\ c & h & l \end{bmatrix}) \quad (29)$$

$$= \begin{bmatrix} (-h^2) & (ch-bl) & (bh) \\ (ch-bl) & (al-c^2) & (bc-ah) \\ (bh) & (bc-ah) & (-b^2) \end{bmatrix} \quad (30)$$

Thus,

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

as requested,

It is notable that FIM should satisfy the symmetry requirement.

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

From (3), the reader can check that

$$\Gamma_{11,1}^{(\alpha)} = -(1 - \alpha) \left(\frac{1}{(1-\rho)^3} + \frac{1}{\rho^3} \right) \quad (31)$$

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

$$\Gamma_{31,1}^{(\alpha)} = (1 - \alpha) \rho \left(\frac{1}{(1-\rho)^3} - \frac{1}{\rho^3} \right) \quad (33)$$

$$\Gamma_{13,1}^{(\alpha)} = \Gamma_{11,3}^{(\alpha)} = (1 - \alpha) \rho \left(\frac{1}{(1-\rho)^3} - \frac{1}{\rho^3} \right) \quad (34)$$

$$\Gamma_{13,2}^{(\alpha)} = \Gamma_{12,3}^{(\alpha)} = \frac{(1-\alpha)}{2} \quad (35)$$

$$\Gamma_{13,3}^{(\alpha)} = \frac{(1-\alpha)}{2} \left[\frac{\rho^*}{\rho^2} + \frac{\rho^*}{(1-\rho)^2} + \frac{2\rho^2}{(1-\rho)^3} + t\mu^* + 2\mu^* \right] \quad (36)$$

$$\Gamma_{23,1}^{(\alpha)} = \Gamma_{21,3}^{(\alpha)} = \frac{(1-\alpha)\rho^*}{2} \quad (37)$$

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

$$\Gamma_{31,2}^{(\alpha)} = \Gamma_{32,1}^{(\alpha)} = \frac{(1-\alpha)}{2} \quad (39)$$

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

4. The IGMES of the Coordinates of the Transient M/M/1 QM

A. The IMEs of the server utilization coordinate, ρ of the transient M/M/1 QM

The IMEs corresponding to the arrival rate coordinate, ρ of the transient M/M/1 QM 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, ρ .

$$\begin{aligned} & \left[\frac{d^2\rho}{dt^2} + \left[\frac{L}{2} + \frac{L\rho^*}{2} \right] \left(\frac{d\rho}{dt} \right) \left(\frac{d\mu}{dt} \right) + \frac{1}{2} \left(\begin{aligned} & 2A\rho \left(\frac{1}{(1-\rho)^3} - \frac{1}{\rho^3} \right) + \\ & B + L \left[\frac{\rho^*}{\rho^2} + \frac{\rho^*}{(1-\rho)^2} + \frac{2\rho^2}{(1-\rho)^3} + t\mu^* + 2\mu^* \right] \right) \right] + \\ & \frac{1}{2} \left(2A\rho \left(\frac{1}{(1-\rho)^3} - \frac{1}{\rho^3} \right) + B + \left(\frac{\rho\rho^*-2\rho^2}{\rho^3} + \frac{2\rho^2+(1-\rho)\rho^*}{(1-\rho)^3} + \right. \right. \\ & \left. \left. t\mu^* + \rho\mu^* + \mu^* + \mu\rho^* \right) L \right) \left(\frac{d\rho}{dt} \right) \left(\frac{dt}{dt} \right) + \left[\rho^* L \left(\frac{1}{(1-\rho)^3} - \frac{1}{\rho^3} \right) \right] \left(\frac{d\rho}{dt} \right)^2 + \\ & \left[\frac{1}{2} \left(A\rho^* + B(1 + 2\rho^* + t\rho^*) \right) + \frac{1}{2} \left(A + (\rho^* + t\rho^*)L \right) \right] \left(\frac{d\mu}{dt} \right) \left(\frac{dt}{dt} \right) + \\ & \left[\frac{1}{2} \left(\left(\frac{\rho\rho^*-2\rho^2}{\rho^3} + \frac{2\rho^2+(1-\rho)\rho^*}{(1-\rho)^3} + t\mu^* + \rho\mu^* + \mu^* + \mu\rho^* \right) A + (\rho^* + t\rho^*)B + \left[\frac{(\rho^*\rho^* - \rho\rho^*)\rho - 2\rho(\rho^2 - \rho\rho^*)}{\rho^3} + \right. \right. \right. \\ & \left. \left. \frac{(\rho^*\rho^* + (1-\rho)\rho^*)(1-\rho) + 2\rho(\rho^2 + (1-\rho)\rho^*)}{(1-\rho)^3} + \right. \right. \\ & \left. \left. (1 + \rho)t\mu^* + (1 + \rho)\mu^* + 3\rho\mu^* + \mu^* + 2\rho\mu^* + 2\mu\rho^* + \right. \right. \\ & \left. \left. 4\mu\rho^* + 2\rho\mu^* + \mu t\rho^* + \mu\rho^* + \mu t\rho^* \right) L \right] \left(\frac{dt}{dt} \right)^2 \right] = 0 \quad (40) \end{aligned}$$

It can be verified that one of the closed form solutions of (40) is determined by the paths of motion:

$$\rho(t) = \text{constant}, \mu(t) = \text{constant}$$

such that

$$\rho \in (0,1), \mu > 1 \quad (41)$$

B. The IMEs of the Mean service Rate coordinate, μ of the transient M/M/1 QM

The IMEs of μ , of the transient M/M/1 QM 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, μ .

$$\begin{aligned} & \left[\frac{d^2\mu}{dt^2} + \left[\frac{F}{2} + \frac{F\rho}{2} \right] \left(\frac{d\rho}{dt} \right) \left(\frac{d\mu}{dt} \right) + \left[\frac{1}{2} \left(2B\rho \left(\frac{1}{(1-\rho)^3} - \frac{1}{\rho^3} \right) + E + F \left[\frac{\rho}{\rho^2} + \frac{\rho}{(1-\rho)^2} + \frac{2\rho^2}{(1-\rho)^3} + t\mu + 2\mu \right] \right) \right. \\ & \quad \left. + \frac{1}{2} \left(2B\rho \left(\frac{1}{(1-\rho)^3} - \frac{1}{\rho^3} \right) + E + \left(\frac{\rho\rho - 2\rho^2}{\rho^3} + \frac{2\rho^2 + (1-\rho)\rho}{(1-\rho)^3} + t\mu \right) \right. \right. \\ & \quad \left. \left. + \rho\mu + \mu + \mu\rho \right) F \right] \left(\frac{d\mu}{dt} \right) \left(\frac{dt}{dt} \right) + \left[\rho \left(\frac{1}{(1-\rho)^3} - \frac{1}{\rho^3} \right) F \right] \left(\frac{d\rho}{dt} \right)^2 + \\ & \quad \left[\frac{1}{2} (B\rho + E(1 + 2\rho + t\rho)) + \frac{1}{2} (B + (\rho + t\rho)F) \right] \left(\frac{d\rho}{dt} \right) \left(\frac{dt}{dt} \right) + \\ & \quad \left[\frac{1}{2} \left(\left(\frac{\rho\rho - 2\rho^2}{\rho^3} + \frac{2\rho^2 + (1-\rho)\rho}{(1-\rho)^3} + t\mu + \rho\mu + \mu + \mu\rho \right) B + \right. \right. \\ & \quad \left. \left. (\rho + t\rho)E + \left[\frac{(\rho\rho - \rho\rho)\rho - 2\rho(\rho^2 - \rho\rho)}{\rho^3} + \frac{(\rho\rho + (1-\rho)\rho)(1-\rho) + 2\rho(\rho^2 + (1-\rho)\rho)}{(1-\rho)^3} \right] \right) \right. \\ & \quad \left. (1 + \rho)t\mu + (1 + \rho)\mu + 3\rho\mu + \mu + 2\rho\mu + 2\mu\rho + 4\mu\rho + 2\rho\mu + \mu t\rho + \mu\rho + \mu t\rho \right] F \left(\frac{dt}{dt} \right)^2 \Big] = 0 \quad (42) \end{aligned}$$

Let $\rho = \text{constant}$, then (42) reduces to

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

$$\Delta = \det([g_{ij}]) =$$

$$\left(\left(\frac{1}{\rho^2} - \frac{1}{(1-\rho)^2} \right) (-((1 + \rho))^2) - (t^2 [((1 + \rho)t\mu + \mu + 2\rho\mu]) \right) \quad (44)$$

$$B = \frac{(t\mu(1 - \rho) - t[((1 + \rho)t\mu + \mu)])}{\Delta} =$$

$$\frac{(t\mu(1 - \rho) - t[((1 + \rho)t\mu + \mu)])}{\left(\left(\frac{1}{\rho^2} - \frac{1}{(1-\rho)^2} \right) (-((1 + \rho))^2) - (t^2 [((1 + \rho)t\mu + \mu + 2\rho\mu]) \right)} \quad (45)$$

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

$$\frac{\left(\left(\frac{1}{\rho^2} - \frac{1}{(1-\rho)^2} \right) [((1 + \rho)t\mu + \mu + 2\rho\mu] - (t\mu)^2 \right)}{\left(\left(\frac{1}{\rho^2} - \frac{1}{(1-\rho)^2} \right) (-((1 + \rho))^2) - (t^2 [((1 + \rho)t\mu + \mu + 2\rho\mu]) \right)} \quad (46)$$

$$F = \frac{(bc-ah)}{\Delta} = \frac{(t(t\mu) - (\frac{1}{\rho^2} - \frac{1}{(1-\rho)^2})(1+\rho))}{\left(\left(\frac{1}{\rho^2} - \frac{1}{(1-\rho)^2}\right) - ((1+\rho)^2) - (t^2[(1+\rho)t\mu + \mu + 2\rho\mu])\right)} \quad (47)$$

Thus,

$$\begin{aligned} & \left[\frac{d^2\mu}{dt^2} + \frac{1}{2} \left(\frac{\left(\frac{1}{\rho^2} - \frac{1}{(1-\rho)^2}\right) [(1+\rho)t\mu + \mu + 2\rho\mu] - (t\mu)^2}{\left(\left(\frac{1}{\rho^2} - \frac{1}{(1-\rho)^2}\right) - ((1+\rho)^2) - (t^2[(1+\rho)t\mu + \mu + 2\rho\mu])\right)} + \right. \right. \\ & \left. \left. \frac{(t(t\mu) - (\frac{1}{\rho^2} - \frac{1}{(1-\rho)^2})(1+\rho))}{\left(\left(\frac{1}{\rho^2} - \frac{1}{(1-\rho)^2}\right) - ((1+\rho)^2) - (t^2[(1+\rho)t\mu + \mu + 2\rho\mu])\right)} [t\mu + 2\mu] \right) + \right. \\ & \left. \frac{\left(t(t\mu) - \left(\frac{1}{\rho^2} - \frac{1}{(1-\rho)^2}\right)(1+\rho)\right)}{\left(\left(\frac{1}{\rho^2} - \frac{1}{(1-\rho)^2}\right) - ((1+\rho)^2) - (t^2[(1+\rho)t\mu + \mu + 2\rho\mu])\right)} + \right. \\ & (t\mu + \mu) \frac{\left(t(t\mu) - \left(\frac{1}{\rho^2} - \frac{1}{(1-\rho)^2}\right)(1+\rho)\right)}{\left(\left(\frac{1}{\rho^2} - \frac{1}{(1-\rho)^2}\right) - ((1+\rho)^2) - (t^2[(1+\rho)t\mu + \mu + 2\rho\mu])\right)} \left] \left(\frac{d\mu}{dt}\right) + \frac{1}{2} ((t\mu + \right. \\ & \left. \mu) \frac{(t\mu(1-\rho) - t[(1+\rho)t\mu + \mu])}{\left(\left(\frac{1}{\rho^2} - \frac{1}{(1-\rho)^2}\right) - ((1+\rho)^2) - (t^2[(1+\rho)t\mu + \mu + 2\rho\mu])\right)} + \right. \\ & \left. [(1+\rho)t\mu + (1+\rho)\mu + \mu + 2\rho\mu] \frac{(t(t\mu) - (\frac{1}{\rho^2} - \frac{1}{(1-\rho)^2})(1+\rho))}{\left(\left(\frac{1}{\rho^2} - \frac{1}{(1-\rho)^2}\right) - ((1+\rho)^2) - (t^2[(1+\rho)t\mu + \mu + 2\rho\mu])\right)} \right] = 0 \quad (48) \end{aligned}$$

It is obvious that for arbitrary constant values of ρ, μ , we have a closed form solution for (48). As time becomes sufficiently large, i.e., $t \rightarrow \infty$, (48) reduces to

$$\left[\frac{3\mu}{2} - \frac{3\mu^2}{2(1+\rho)} - \frac{\mu\mu}{\mu} \right] = 0 \quad (49)$$

We propose the closed form solution, $\mu = \frac{2\gamma t}{(1+\gamma t)}$, for some arbitrary non-zero constant γ . substituting in (49) implies:

$$18\gamma^3 + \frac{6\gamma^2}{(1+\gamma t)(1+\rho)} + 6\gamma^4 = 0 \quad (50)$$

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 $t \rightarrow \infty$.

C. The IMEs of the temporal coordinate, t of the transient M/M/1 QM

Engaging a similar lengthy mathematical manipulation, it can be verified that the IMEs of the temporal coordinate, t of the transient M/M/1 QM, as time reaches infinity and for a constant server utilization will have a closed form solution that is characterized by the family of families of temporal curves:

$$\mu(t) = \zeta_1 + \zeta_2 t + \zeta_3 t^2 \quad (51)$$

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

5. 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 a parameter space of probability distributions. Amari[3] 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[11].

In reinforcement learning(RL) [12], 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[12] 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 reinforcement learning.

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 Figure 2.

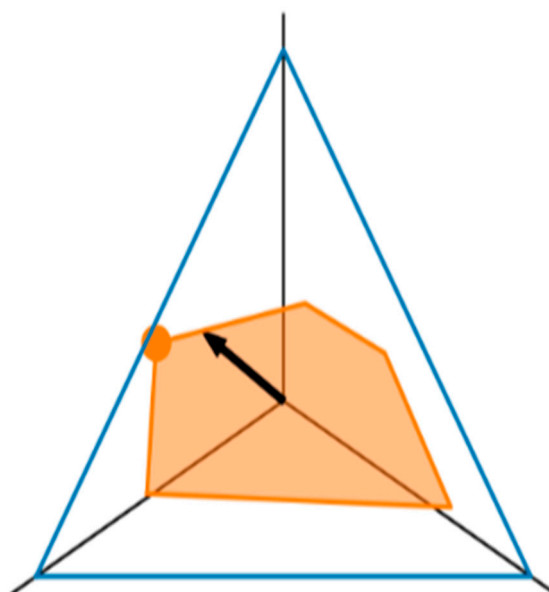


Figure 2. We can represent reward functions as vectors starting from the origin. The objective of maximizing expected return is equivalent to maximizing the dot product between the state marginal distribution and the reward vector. This visualization helps us understand how different reward-maximizing policies are related to the state marginal polytope and provides insights into the relationship between state-dependent reward functions and the optimal policies[12].

The analysis of where the skills learned by Mutual Information-based Skill Learning (MISL) are positioned on the state marginal polytope, which has motivated [12] 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[12], 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 that are as distinct as possible from each other and from 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 state marginals that are equidistant 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 by Figure 3.

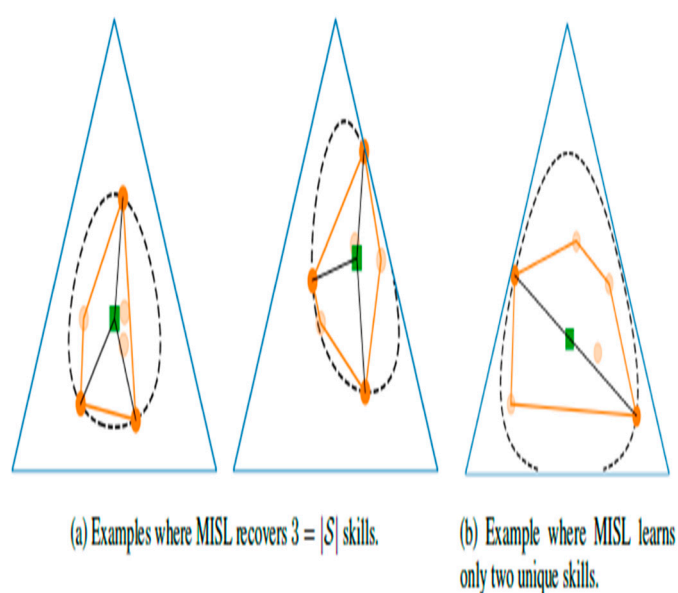


Figure 3. Skill Learning's IG [12].

Reliable detection of out-of-distribution (OOD) samples[13] is crucial for ensuring the safety of modern machine learning systems. The work of [13] 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[13].

The authors [13] 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.

Figure 4 showcases a comparison between Fisher-Rao and Mahalanobis distances for distinguishing between 1D Gaussian distributions in the context of out-of-distribution (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.

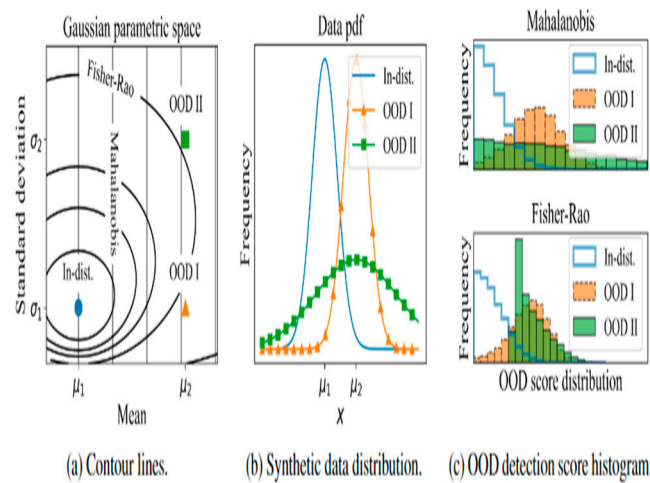


Figure 4. Portraying the reason behind using Fisher-Rao OOD detection[13].

In the context of the experiment, the authors [13] compared three different settings for out-of-distribution (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 Figure 5.

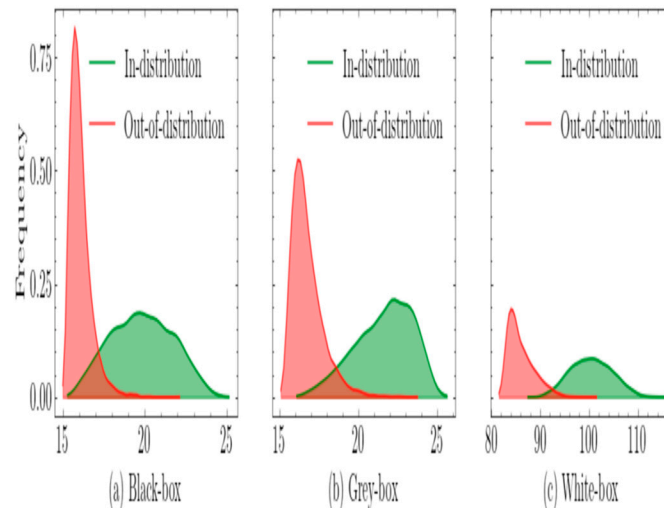


Figure 5. The probability distributions of the IGEOOD score, which is a metric used for out-of-distribution (OOD) detection. The distributions are shown for three different settings, specifically for a pre-trained DenseNet model on CIFAR-10 dataset, considering both in-distribution and OOD data from TinyImageNet (downsampled). This analysis helps evaluate the effectiveness of the IGEOOD score in distinguishing between in-distribution and OOD samples[13].

6. Conclusion, Open Problems and Future Work

This study offers a revolutionary info-geometrics of transient $M/M/1$ QM. For this queue, FIM and IFIM are established. The geodesic equations 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:

According to the analysis in [12], 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

$$\Psi(\theta) = -\ln(\rho) - \ln(1 - \rho) + \mu(1 + \rho)t, p_0(t) = 1 - \rho \text{ (c.f., (21))}$$

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 (c.f., (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.

References

1. A. Mageed, Y. Zhou, Y. Liu and Q. Zhang, "Towards a Revolutionary Info-Geometric Control Theory with Potential Applications of Fokker Planck Kolmogorov(FPK) Equation to System Control, Modelling and Simulation," 2023 28th International Conference on Automation and Computing (ICAC), Birmingham, United Kingdom, 2023, pp.1-6, doi: 10.1109/ICAC57885.2023.10275271
2. Z. Skoda, Z., 2019, available online at <https://ncatlab.org/nlab/show/information+geometry>
3. A. Mageed, Q. Zhang, T. C. Akinci, M. Yilmaz and M. S. Sidhu, "Towards Abel Prize: The Generalized Brownian Motion Manifold's Fisher Information Matrix With Info-Geometric Applications to Energy Works," 2022 Global Energy Conference (GEC), Batman, Turkey, 2022, pp. 379-384, doi: 10.1109/GEC55014.2022.9987239.
4. A. Mageed, X. Yin, Y. Liu and Q. Zhang, " $Z_{(a,b)}$ 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," 2023 28th International Conference on Automation and Computing (ICAC), Birmingham, United Kingdom, 2023, pp. 1-6, doi: 10.1109/ICAC57885.2023.10275185.
5. K. Nakagawa, " The geometry of $M/D/1$ queues on Large Deviation," International Transactions of Operational Research, vol. 9, 2002.
6. C.T. J. Dodson, "Topics in Information Geometry," University of Manchester, 2005, Available online on <http://eprints.maths.manchester.ac.uk/131/1/InfoGeom.pdf>
7. L. Peg, H. Sun, and J. Jiu, 2007, "The Geometric Structure of the Pareto Distribution, " Boletín de la Asociación Matemática Venezolana, vol . XIV, no.1, 2007.

8. A. Mageed and D. D. Kouvatso, "Information Geometric Structure of Stable M/G/1 Queue Manifold and its Matrix Exponential," Proceedings of the 35th UK Performance Engineering Workshop, School of Computing, University of Leeds, Edited by Karim Djemame, 2019, p. 123-135. [Online] at: <https://sites.google.com/view/ukpew2019/home>
9. I.A.Mageed and D. D. Kouvatso, " The Impact of Information Geometry on the Analysis of the Stable M/G/1 Queue Manifold," In Proceedings of the 10th International Conference on Operations Research and Enterprise Systems - ICORES, ISBN 978-989-758- 485-5, vol. 1, 2021, p. 153-160. DOI: 10.5220/0010206801530160.
10. [10] P. R. Parthasarathy, "A transient solution to an M /M /1 queue: a simple approach," Adv. Appl. Prob., vol. 19, 1987, p. 997-998.
11. I. A. Mageed and Q. Zhang, " Information Geometry? Exercises de Styles," electronic Journal of Computer Science and Information Technology, vol. 8, no. 1, 2022, p. 9-14.
12. B. Eysenbach, R. Salakhutdinov, and S. Levine S, " The information geometry of unsupervised reinforcement learning," arXiv preprint arXiv:2110.02719, 2021.
13. E. D. Gomes, et al, " Igeood: An information geometry approach to out-of-distribution detection," arXiv preprint arXiv:2203.07798,2022.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.