

## **A Note on a Possible Connection between Pontryagin Maximum Principle and Jaynes Maximum Entropy**

*Javier Rivera*

865 River Boat Circle, Orlando, FL 32828

Copyright © 2018 USA Registration # TXu 1-878-887

Email: [caribe31415@gmail.com](mailto:caribe31415@gmail.com)

*Received 6 May 2018; accepted 26 June 2018*

**Abstract.** A conjecture on the possible connection between Pontryagin Maximum Principle and Jaynes Maximum Entropy has been developed in this article. A known manufacturing facility or plant wants to optimally reduce the sum of the variances of its observed production process. The plant's models consist of known controllable state models along with known observational models of production. The selection of the plants control vector can be described as an optimal control problem. Therefore, the Pontryagin Maximum Principle will be used to discover a possible connection to Jaynes Maximum Entropy. The conjectured connection of Pontryagin Maximum Principle to Jaynes Maximum Entropy can be derived from each other under certain transformations relating the probability space and conjugate vector space as defined in a Pontryagin Hamiltonian system with the probability space as defined in Shannon's entropy. The conjugate vector space is used to define a known set of probabilities, which associates the Hamiltonian of the system to Jaynes Maximum Entropy via the Legendre transformation.

**Keywords:** Pontryagin maximum principle, Jaynes maximum entropy, Legendre transformation

**AMS Mathematics Subject Classification (2010):** 34H05, 34K35

### **1. Introduction**

The articles of Wang derived the relationship between maximum entropy change [1] and maximum path information [2] with the Hamilton principle of least action for non-equilibrium systems. Given Wang results for maximum entropy implies there is reasonable conjecture for a possible connection of the Pontryagin Maximum Principle to the Jaynes Maximum Entropy since the Pontryagin Maximum Principle is an extension of Hamilton principle of least action [3].

In this article, the plant models are based mostly on statistical predictor-corrector models that can be written as known state and observation nonlinear models. Consider an example: a known plant wants to reduce errors in production by minimizing the sums of observed production variances through time. The description of the steps needed to treat this nonlinear optimal control process with the Pontryagin Maximum Principle

needs the definition of the plant states nonlinear models and the plant sensors observation given by known nonlinear models. This problem will not be solved but rather the aim of this article is to conjecture a connection between the Pontryagin Maximum Principle and Jaynes Maximum Principle as a mapping developed between the probability space and the conjugate vector space. The Legendre transformation is used to map the Pontryagin Hamiltonian of the maximum principle to a Pontryagin Lagrangian. It follows the Pontryagin Lagrangian is just the definition of entropy as used by Jaynes Maximum Entropy. The definition of a plant model will be as found in Smirnov's book "*Oscillation Theory of Optimal Process*" [4].

## 2. Known plant models and plant sensors

A manufacturing plant wants to reduce errors in production by minimizing the sums of variances through time subject to the known state and observation model's constraints. The following nonlinear system of differential equations characterizes the known plant state models (Eq. 1).

$$\dot{\underline{x}}_n(t) = \underline{f}_n(\underline{x}(t), \underline{u}_n(t)) \quad n = 1 \text{ to } N; \quad (1)$$

Note  $\underline{x}(t) = (\underline{p}_1^T(t), \underline{\dot{p}}_1^T(t), \underline{p}_2^T(t), \underline{\dot{p}}_2^T(t), \dots, \underline{p}_n^T(t), \underline{\dot{p}}_n^T(t))^T$  and

$\underline{x}_n(t) = (\underline{p}_n^T(t), \underline{\dot{p}}_n^T(t))^T$ , where T denotes transpose operator;  $\underline{p}_n^T(t)$ , and  $\underline{\dot{p}}_n^T(t)$  represent the  $n^{\text{th}}$  plant state 3-component position and 3 component velocity row vectors.  $\underline{u}_n(t)$  is the  $n^{\text{th}}$  plant parametric feature control column vector. The function  $\underline{f}_n$  is the  $n^{\text{th}}$  plant state's known nonlinear model. The sensors,  $\underline{S}_k$ , observes the  $N$  plants states,  $\underline{x}(t)$ , with some additive environmental noise,  $\underline{w}_k(t)$ , with a known probability density function is given by (Eq. 2).

$$\underline{Z}_k(t) = \underline{S}_k(\underline{x}(t), \underline{u}(t)) + \underline{w}_k(t) \quad \text{where } k = 1 \text{ to } K, \quad K > N; \quad (2)$$

$\underline{Z}_k(t) = (\underline{z}_k^T(t), \underline{\dot{z}}_k^T(t))^T = Z_{i,k}(t)$  where index  $i$  equal 1 to 3 are the  $k^{\text{th}}$  measured plant's production position coordinates, and index  $i$  equal 4 to 6 are the observed  $k^{\text{th}}$  measurement plant's production rate or "velocity" vector.  $\underline{z}_k^T(t)$  is the observed  $k^{\text{th}}$  measurement plant's production "position" vector with respect to the common origin with the same dimensions as the state vector,  $\underline{x}_n(t)$ .  $\underline{\dot{z}}_k^T(t)$  is the observed  $k^{\text{th}}$  measurement plant's production rate or "velocity" vector.  $\underline{x}(t)$  is the state vector representing  $N$  plant states relative a common origin.  $\underline{u}(t)$  is the combined or aggregate known feature control parameters vectors.  $\underline{w}_k(t)$  is the  $k^{\text{th}}$  additive noise position and velocity with zero mean process due to the plant sensor limitations, and other the environmental conditions. Thus  $\underline{S}_k$  is a nonlinear sensor measurement function based on the combined plant state vectors, and combined control parameters. In this example, the purpose of this control process is to move the initial plant state from  $\underline{x}(t_0) = \underline{x}_0$  to a final state  $\underline{x}(t_1) = \underline{x}_1$  with minimum observational model variance. The Pontryagin Maximum Principle method a functional is given a priori, to be maximized, typically this functional is given a negative absolute value therefore the maximum is known to be zero. The following four steps are the usual "set up" steps of a dynamical system (Eqs. 1 and 2) to be solved with Pontryagin Maximum Principle [3]:

1. Differentiate (Eq. 2) the observation models with respect to time *without* the differentiation of the control parameter vector,  $\underline{u}$ , to obtain (Eq. 3A or 3BC)

A Note on a Possible Connection between Pontryagin Maximum Principle and Jaynes Maximum Entropy

$$\frac{dZ_{i,k}(t)}{dt} = \sum_{n=1}^N \sum_{j=1}^6 f_{j,n}(\underline{x}(t), \underline{u}(t)) \frac{\partial}{\partial x_{j,n}} [S_{i,k}(\underline{x}(t), \underline{u}(t))]_{\underline{u}} + \frac{dw_{i,k}}{dt} \quad (3A)$$

Or

$$\frac{dZ_{i,k}(t)}{dt} = G_{i,k}(\underline{x}(t), \underline{u}(t)) + \frac{dw_{i,k}}{dt} \quad (3B)$$

With

$$G_{i,k}(\underline{x}(t), \underline{u}(t)) = \sum_{n=1}^M \sum_{j=1}^6 f_{j,n}(\underline{x}(t), \underline{u}(t)) \frac{\partial}{\partial x_{j,n}} [S_{i,k}(\underline{x}(t), \underline{u}(t))]_{\underline{u}}$$

2. Rename the observation variables, and the hidden state variables into a new expanded state vector  $\underline{Y}(t)$ , (Eq. 4).

$$\underline{Y}(t) = (\underline{x}^T(t), \underline{Z}^T(t))^T \quad (4)$$

where  $\underline{Z}(t)$  vector is the aggregate of the  $K$  sensors observations  $(\underline{Z}_1^T(t), \dots, \underline{Z}_K^T(t))^T$ .

3. Define the minimum error variance,  $\sigma^2$ , in (Eq. 5) as the process for obtaining an optimal control which minimizing the sum of observational noise variance defined by following equation assuming *zero mean noise process*,  $E\{\underline{w}_k\} = 0$ .

$$\sigma^2 = \sum_{k=1}^K E\{(\underline{w}_k)^2\} = \int_{-\infty}^{\infty} \sum_{k=1}^K \sum_{l=1}^6 (w_{i,k})^2 D(w_{i,k}) dw_{i,k} \quad (5)$$

In (Eq. 5),  $D(w_{i,k})$ , are known probability density function of the  $w_{i,k}$  observation noise processes. Since the noise process is a zero mean process the measurements of state are unbiased.

Define the sensor  $k$  to be  $\underline{S}_k(\underline{x}(t), \underline{u}(t))$  into a vector composed of the following combined series of sensors

components  $\{S_{6N+6k-5}(\underline{Y}(t), \underline{u}(t)), \dots, S_{6N+6k}(\underline{Y}(t), \underline{u}(t))\}$  for  $k = 1$  to  $K$ .

$$\underline{S} = (S_{6N+6}, \dots, S_{6N+6}, \dots, S_{6N+6k-5}, \dots, S_{6N+6k}, \dots, S_{6N+6k-5}, \dots, S_{6N+6k})$$

Similarly define  $\underline{G}_k(\underline{x}(t), \underline{u}(t))$  into a vector of aggregate or combined series of components

$$\underline{G} = (G_{6N+6}, \dots, G_{6N+6}, \dots, G_{6N+6k-5}, \dots, G_{6N+6k}, \dots, G_{6N+6k-5}, \dots, G_{6N+6k})$$

The maximization of a functional,  $J[\underline{Y}(\tau), \underline{u}(\tau)]$ , in terms of the new state variable  $\underline{Y}$  and feature enhanced control parameters  $\underline{u}$  by a change of variables,  $dw_{i,k} = \frac{dw_{i,k}}{d\tau} d\tau$  is given by (Eq. 6)

$$J = J[\underline{Y}(\tau), \underline{u}(\tau)] = - \int_{t_0}^{t_1} |I(\underline{Y}(\tau), \underline{u}(\tau))| d\tau \leq 0 \quad (6)$$

Note that the maximum of  $J$  is zero, where the integrand,  $I(\underline{Y}, \underline{u})$ , is given by (Eq. 7).

$$I(\underline{Y}, \underline{u}) = \sum_{k=1}^{6K+6N} (Y_k - S_{k-6N}(\underline{Y}, \underline{u}))^2 D((Y_k - S_{k-6N}(\underline{Y}, \underline{u}))(\frac{dY_k}{d\tau} - G_{k-6N}(\underline{Y}, \underline{u}))) \quad (7)$$

Subject to the state model system of differential equations:

$$\frac{dY_r}{dt} = F_r(\underline{Y}(t), \underline{u}(t)) \quad \text{where } r = 1 \text{ to } 6N + 6K. \quad (8)$$

where  $\underline{F}$  is given as the following vector  $(f_1, \dots, f_N, \underline{G}_1 + \frac{dw_1}{dt}, \dots, \underline{G}_K + \frac{dw_K}{dt})$  of length  $6N+6K$ . Note the control process is to move the initial plant state from

Javier Rivera

$\underline{Y}(t_0) = (\underline{x}^T(t_0), \underline{z}^T(t_0))^T$  to a final state  $\underline{Y}(t_1) = (\underline{x}^T(t_1), \underline{z}^T(t_1))^T$  with minimum sum of observational models variances, yet other cost functions could be used such as minimum time. An equivalent method would be to instead use Lagrange multipliers [3] yet the method used here is simpler since the Lagrange multipliers introduces additional unknowns.

4. Define an additional variable

$$Y_0(t) = - \int_{t_0}^t |I(\underline{Y}(\tau), \underline{u}(\tau))| d\tau \quad (9)$$

By differentiation of (Eq. 9), we obtain (Eq. 10).

$$\frac{dY_0(t)}{dt} = -|I(\underline{Y}(t), \underline{u}(t))| = F_0(\underline{Y}(t), \underline{u}(t)) \quad (10)$$

Note that the initial conditions for (Eq. 10) are  $Y_0(t_0) = 0, Y_0(t_1) = J$ .

For the Pontryagin Maximum Principle definitions and methods please refer to L.S. Pontryagin et al, “*The Mathematical Theory of Optimal Process*” pages79-80 [3] or Smirnov’s book entitled “*Oscillation Theory of Optimal Processes*” pages 7-8 [4].

The Pontryagin Maximum Principle converts a very difficult observational/control problem to an operations research problem in which we know the optimal results are at the edges or vertices of the boundary that intercept of the control parameter volume by the “cost” or Hamiltonian function. The Pontryagin Maximum Principle finds the minimum observational noise variance given the known plant states and observation models as constraints; Pontryagin Maximum Principle maps the control process problem into an Operations Research problem [3]. The Pontryagin Hamiltonian is defined with the help of the conjugate vector space,  $\Psi_\alpha(t)$ , in (Eq. 11) [3].

$$H(\underline{\Psi}(t), \underline{Y}(t), \underline{u}(t)) = \sum_{\alpha=0}^{6N+6K} \Psi_\alpha(t) F_\alpha(\underline{Y}(t), \underline{u}(t)) \quad (11)$$

where the equality holds almost everywhere. The Pontryagin Hamilton equations [3] are given by (Eqs. 12AB).

$$\frac{dY_r}{dt} = \frac{\partial H}{\partial \Psi_r} = F_r(\underline{Y}(t), \underline{u}(t)) \quad (12A)$$

$$\frac{d\Psi_r}{dt} = - \frac{\partial H}{\partial Y_r} = - \sum_{\alpha=0}^{6N+6K} \Psi_\alpha(t) \frac{\partial F_\alpha(\underline{Y}(t), \underline{u}(t))}{\partial Y_r} \quad (12B)$$

Index r is within the integers 0 to  $6N + 6K$ . The characteristic functions  $F_\alpha(\underline{Y}(t), \underline{u}(t))$  will hopefully be sparse in the functional dependence on the state variable,  $Y_r$ , therefore the matrix  $F_{\alpha,r} = \frac{\partial F_\alpha(\underline{Y}(t), \underline{u}(t))}{\partial Y_r}$  will be sparse. Using the usual calculus necessary, and sufficient conditions to perform the maximization of the Hamiltonian.

The least upper bound value or supremum of the Pontryagin Hamiltonian,

$$M(\underline{Y}(t), \underline{u}(t)) = \sup_{\underline{u} \in V} H(\underline{\Psi}(t), \underline{Y}(t), \underline{u})$$

typically lies in the boundary of the control space. There are numerical methods such as steepest descent which could be used to find the optimal observational control parameters  $\underline{u}(t)$  to move the initial plant state to a final state with minimum observational model variance. But this is not the aim of this article, but to conjecture a possible connection of Pontryagin Maximum Principle with Jaynes Maximum Entropy.

**2. Conjecture: Jaynes maximum entropy can be derived from Pontryagin maximum principle**

We can conjecture the relationship between Pontryagin Maximum Principle and Jaynes Maximum Entropy can be achieved by mapping of the conjugate vector,  $\Psi_\alpha$ , to a probabilistic space,  $E_\alpha$ . The exact definition of the probabilistic mapping is not defined in this article since its part of the conjecture that at least one such mapping exists.

Given random *disjoint* events,  $E_\alpha$ , in a known probability space, where  $S = \bigcup_{\alpha=0}^{6N+6K} E_\alpha$  and the probabilities,  $P_\alpha(t)$ , can be defined by a conjectured probability mapping of the events,  $E_\alpha$ , by a function  $\varphi$  of the conjugate vector and finite norm,  $\rho(t)$ , of the conjugate vector:

$$\rho(t) = \|\underline{\Psi}(t)\| \gg 1, t \text{ in } [t_0, t_1]$$

$$P_\alpha(t) = P(E_\alpha, t) = \varphi_\alpha(\rho(t), \underline{\Psi}(t)) \geq 0$$

The probability of event,  $E_\alpha$ , due to its disjoint property of the members is given by

$$P(S) = \sum_{\alpha=0}^{6N+6K} P(E_\alpha, t) = \sum_{\alpha=0}^{6N+6K} P_\alpha(t) = 1$$

The above constraint equation plays the role as the sum of probabilities of all events that yields to one. An example of a probability mapping,  $\varphi_\alpha$ , is the exponential mapping found in [1,2].

Then by setting  $P_\alpha(t)$  as a conjugate vector into a new Hamiltonian formulation could be defined as an expectation of  $F_\alpha(\underline{Y}(t), \underline{u}(t))$ .

$$\tilde{H}(\underline{P}(t), \underline{Y}(t), \underline{u}(t)) = \sum_{\alpha=0}^{6N+6K} P_\alpha(t) F_\alpha(\underline{Y}(t), \underline{u}(t)) = E\{F\}.$$

With the constraint

$$\sum_{\alpha=0}^{6N+6K} P_\alpha(t) = 1$$

Since the following relation holds for all conjugate vector and conjugate probabilities by inspection.

$$\tilde{H}(\underline{P}(t), \underline{Y}(t), \underline{u}(t)) \geq H\left(\frac{\underline{\Psi}(t)}{\rho(t)}, \underline{Y}(t), \underline{u}(t)\right) = \frac{1}{\rho(t)} H(\underline{\Psi}(t), \underline{Y}(t), \underline{u}(t)).$$

Then the least upper bound of the Hamiltonian as a function of the control parameter  $\underline{u}$  holding the state vector,  $\underline{Y}(t)$ , and conjugate vector constant,  $\underline{\Psi}(t)$ , therefore  $\rho(t)$  and the probability of state,  $P_\alpha$ , are also held constant then

$$\begin{aligned} \tilde{M}(\underline{Y}(t), \underline{u}(t)) &= \sup_{\underline{u} \in V} \rho(t) \tilde{H}(\underline{P}(t), \underline{Y}(t), \underline{u}) \\ &\geq M(\underline{Y}(t), \underline{u}(t)) = \sup_{\underline{u} \in V} H(\underline{\Psi}(t), \underline{Y}(t), \underline{u}) \end{aligned}$$

Since the Hamiltonian is defined as a linear function with respect to the conjugate vector in the Pontryagin Maximum Principle, note that the Maximum Principle may be considered the negative of the Hamiltonian used in Hamilton's Least Action (with unit mass, i.e. probability space). Now a question arises: What the Pontryagin Lagrangian

can be functionally defined? In Theorem 1, with the help of the Legendre transformation [7] the Pontryagin Lagrangian can be defined in terms of the Pontryagin Hamiltonian.

**Theorem 1. Pontryagin Lagrangian**

If the conjugate probabilities are well defined, and the Legendre transformation (Eq. 13) of the Pontryagin Lagrangian is a valid transformation then the Legendre transformation can be re-interpreted as a partial differential equation of the Pontryagin Lagrangian.

$$-\tilde{H} = \sum_{\alpha=0}^{6N+6K} \left\{ P_{\alpha}(t) \frac{\partial \tilde{L}}{\partial P_{\alpha}(t)} \right\} - \tilde{L} \quad (13)$$

**Proof:** Assuming the conjugate probabilities are well defined and we have a valid Legendre transformation of the conjugate probabilities then we can use the Legendre transformation (Eq. 13) to find the equivalent Pontryagin “Lagrangian” function,  $\tilde{L}$ , where we re-interpret the Legendre transformation (Eq. 13) as a valid partial differential equation with respect to the conjugate probabilities,  $P_{\alpha}(t)$ , coordinates.

$$\sum_{\alpha=0}^{6N+6K} \left\{ P_{\alpha}(t) \frac{\partial \tilde{L}}{\partial P_{\alpha}(t)} \right\} - \tilde{L} = - \sum_{\alpha=0}^{6N+6K} P_{\alpha}(t) F_{\alpha}(\underline{Y}(t), \underline{u}(t)) = -\tilde{H}$$

This equation can be solved for a particular solution of the Pontryagin *Lagrangian* by induction [8] and neglect the homogeneous solutions.

*Case 0:*

$$P_0(t) \frac{d\tilde{L}}{dP_0(t)} - \tilde{L} = -P_0(t) F_0(\underline{Y}(t), \underline{u}(t))$$

This is a linear differential equation of the Pontryagin Lagrangian,  $\tilde{L}$ , its elementary solution is

$$\tilde{L} = -F_0(\underline{Y}(t), \underline{u}(t)) P_0(t) \text{Log}(P_0(t))$$

The *Log* function is the natural logarithm function to the base *e*.

Assume *Case n* holds, that is  $\tilde{L} = \sum_{i=0}^n -F_i(\underline{Y}(t), \underline{u}(t)) P_i(t) \text{Log}(P_i(t))$  is a particular solution of the Legendre transformation (PDE). The case *n+1* needs to be proven as follows.

*Case n+1:*

$$\sum_{\alpha=0}^{n+1} \left\{ P_{\alpha}(t) \frac{\partial \tilde{L}}{\partial P_{\alpha}(t)} \right\} - \tilde{L} = - \sum_{\alpha=0}^{n+1} P_{\alpha}(t) F_{\alpha}(\underline{Y}(t), \underline{u}(t)) = -\tilde{H} = -E\{F\}$$

By construction

$$\tilde{L} = - \sum_{i=0}^{n+1} F_i(\underline{Y}(t), \underline{u}(t)) P_i(t) \text{Log}(P_i(t))$$

Differentiate the Pontryagin Lagrangian with respect to the known probability to obtain

$$\frac{\partial \tilde{L}}{\partial P_{\alpha}(t)} = -F_{\alpha}(\underline{Y}(t), \underline{u}(t)) \{1 + \text{Log}(P_{\alpha}(t))\}$$

Substituting into the Legendre transformation proves the *n+1* case.

A Note on a Possible Connection between Pontryagin Maximum Principle and  
Jaynes Maximum Entropy

$$\begin{aligned}
 \sum_{\alpha=0}^{n+1} \{P_{\alpha}(t) \frac{\partial \tilde{L}}{\partial P_{\alpha}(t)}\} - \tilde{L} &= \\
 &= \sum_{\alpha=0}^{n+1} \left\{ -P_{\alpha}(t) F_{\alpha}(\underline{Y}(t), \underline{u}(t)) \{1 + \text{Log}(P_{\alpha}(t))\} \right\} \\
 &+ \sum_{\alpha=0}^{n+1} F_{\alpha}(\underline{Y}(t), \underline{u}(t)) P_{\alpha}(t) \text{Log}(P_{\alpha}(t)) = - \sum_{\alpha=0}^{n+1} P_{\alpha}(t) F_{\alpha}(\underline{Y}(t), \underline{u}(t)) \\
 &= -E\{\underline{F}\} = -\tilde{H}
 \end{aligned}$$

Therefore, we have shown it holds for all integers n. Therefore, the Pontryagin Lagrangian has the form

$$\tilde{L} = - \sum_{i=0}^{6N+6K} F_i(\underline{Y}(t), \underline{u}(t)) P_i(t) \text{Log}(P_i(t))$$

The time integral of the Pontryagin Lagrangian is given as

$$\begin{aligned}
 J_L &= \int_{t_0}^{t_1} \tilde{L} dt = \int_{t_0}^{t_1} \sum_{i=0}^{6N+6K} -F_i(\underline{Y}(t), \underline{u}(t)) P_i(t) \text{Log}(P_i(t)) dt \\
 &= \int_{-\infty}^{\infty} \sum_{i=0}^{6N+6K} -P_i \text{Log}(P_i) dY_i
 \end{aligned}$$

due to the following change of integration variables,  $F_i dt = \frac{dY_i}{dt} dt = dY_i$ . Replace  $F_i$  with its equivalent vector  $(\frac{dY_0}{dt}, \frac{dx_1}{dt}, \dots, \frac{dx_N}{dt}, \frac{dz_1}{dt}, \dots, \frac{dz_K}{dt})$  of length  $6N+6K+1$ .

$$\begin{aligned}
 J_L &= \int_{t_0}^{t_1} \sum_{i=1}^{6N} -P_i(t) \text{Log}(P_i(t)) \frac{dx_i}{dt} dt \\
 &+ \int_{t_0}^{t_1} \sum_{i=6N+1}^{6N+6K} -P_i(t) \text{Log}(P_i(t)) \frac{dz_{i-6N}}{dt} dt \\
 &+ \int_{t_0}^{t_1} -P_0(t) \text{Log}(P_0(t)) \frac{dY_0}{dt} dt
 \end{aligned}$$

With a change of variables ( $t = f^{-1}(x_i(t))$ , etc.) the action integral can be seen to be as the maximization of entropy in terms of state or phase space, observational probability space, and for the sum of the variances in the observation space.

$$J_L = \int_{-\infty}^{\infty} \sum_{i=1}^{6N} -P_i \text{Log}(P_i) dx_i + \int_{-\infty}^{\infty} \sum_{i=6N+1}^{6N+6K} -P_i \text{Log}(P_i) dz_{i-6N} + \int_{-\infty}^{\infty} -P_0 \text{Log}(P_0) dY_0$$

Since the above mathematical method is reversible, thereby the reverse of above steps could be shown: the maximization of entropy of above action functional of the Pontryagin Lagrangian below

Javier Rivera

$$\tilde{L} = \sum_{i=0}^{6N+6K} -F_i(\underline{Y}(t), \underline{u}(t)) P_i(t) \text{Log}(P_i(t))$$

is equivalent to the Pontryagin Maximum Principle with the use of the Legendre Transformation. Additionally, any other constraints can be dealt with Lagrange multipliers.

For methods of Jaynes Maximum Entropy please refer to his 1957 and 1982 papers [5,6].

**Acknowledgements.** The author would like to thank the anonymous reviewers and chief editor of APAM for their time and effort spent in reviewing this article.

### 3. Conclusion

The conjecture of a connection of Pontryagin Maximum Principle to Jaynes Maximum Entropy can be derived from each other under certain transformations relating the phase space conjugate vector as defined in a Hamiltonian system with the probability space as defined in Shannon's entropy. This connection between the Pontryagin Maximum Principle and Jaynes Maximum Entropy was shown in this note for a manufacturing facility or plant wants to optimally reduce the sum of the variances of its observed production process. The conjugate vector was used to define a uniform set of probabilities, which associates the Hamiltonian of the system to Jaynes Maximum Entropy via the Legendre transformation.

### REFERENCES

1. Q.A.Wang, Maximum entropy change and least action principle for nonequilibrium systems, *Astrophysics and Space Science*, 305(3) (2004) 273-281.
2. Q.A.Wang, Maximum path information and the principle of least action for chaotic system, *Chaos, Solitons and Fractals*, 23(4) (2004) 1253-1258.
3. L.S.Pontryagin, V.G.Boltyanskii, R.V.Gamkrelidze, E.F.Mishchenko, *The Mathematical Theory of Optimal Process*, L.S. Pontryagin Selected Works Volume 4, Gordon and Breach Science Publishers, NY 1986.
4. G.M.Smirnov, *Oscillation Theory of Optimal Process*, Wiley Inter-Science Publication, John Wiley and Sons, NY 1984
5. E.T.Jaynes, Information Theory and Statistical Mechanics, *Physical Review*, 106(4) 1957 (1957) 620-630.
6. E.T.Jaynes, On the rationale of maximum-entropy methods, *Proceedings of the IEEE*, 70(9) (1982) 939-952.
7. I.M.Gelfand and S.V.Fomin, *Calculus of Variations*, Prentice Hall, NJ, (1963).
8. G.E.Andrews, *Number Theory*, Dover Publications, New York, (1971).