Is Robustness of Stochastic Uncertain Systems Related to Information Theory and Statistical Mechanics?

Charalambos D. Charalambous[a] and Andreas Kyprianou[b]

[a]Department of Electrical and Computer Engineering;
[b]Department of Mechanical and Manufacturing Engineering
University of Cyprus, Nicosia, Cyprus

ABSTRACT

Fifty years ago, when Claude Shannon was developing the Mathematical Theory of Communications, for reliable data transmission, which evolved into the subject of information theory, another discipline was developing dealing with Feedback Control of Dynamical System, which evolved into a scientific subject dealing with decision, stability, and optimization. More recently, a separate discipline dealing with robustness of uncertain systems was born in response to the codification of high performance and reliability in the presence of modeling uncertainties. In principle, robustness in dynamical systems is captured through power dissipation via induced norms and dynamic games, while reliable data transmission is captured through measures of information via entropy, relative entropy, and certain laws of Large Deviations theory. The main ingredient in Large Deviations is the rate functional (or action functional in the classical mechanics terminology), often identified through the Cramer or Legendre-Fenchel Transform. On the other hand, robustness of stochastic uncertain systems is currently under development, using information theoretic as well as statistical mechanics concepts, such as, partition functions, free energy, relative entropy, and entropy rate functional.
This lecture will summarize certain connections between fundamental concepts of robustness, information theory, and statistical mechanics, and possibly make future projections into the convergence of these disciplines.

Keywords: Stochastic Systems, Robustness, Information Theory, Statistical Mechanics, Games, Large Deviations

1. INTRODUCTION

In Statistical Mechanics, the basic concept is the partition function, which describes the various states of a system being in an equilibrium. Once the partition function is computed all thermodynamic properties of the system are identified.
In Robust Control, the basic mathematical quantity is the H^∞-norm or Induced Norm and the dissipation inequalities, which state that the output power of the system is less than or equal to the input power of the system.
In Information Theory, the fundamental concept is the entropy of a Random Variable and the Entropy Rate of a sequence of Random Variables, which measure the amount of information or uncertainty associated with the underlying random experiment.
In Large Deviations Theory, the basic concept is the Law of Large Numbers, which allows the computation of the rate at which the probabilities of certain events decay to zero, exponentially fast. This decay is determined once the rate functional is identified.
The scope of this paper to provide a review of recent results found in,[11, 19, 20] and to introduce additional notions, which connect Large Deviations Theory to Robustness and Dissipation. The first result of this paper is to give a variational interpretation of the basic principle of statistical mechanics, which is subsequently employed to relate

Email: chadcha@ucy.ac.cy, Telephone: +357 22 892253, Fax: +357 22892260
C. D. Charalambous is also with the School of Information Technology and Engineering, University of Ottawa, 161 Louis Pasteur, A519, Ottawa, Ontario K1N 6N5, Canada.
This work is supported by the European Commission under the project ICCCSYSTEMS and by the University of Cyprus via a 2004 Medium Size Research Grant.

Noise in Communication Systems, edited by Costas N. Georghiades,
Langford B. White, Proc. of SPIE Vol. 5847 (SPIE, Bellingham, WA, 2005)
0277-786X/05/$15 · doi: 10.1117/12.611785

the Partition Function, the Free Energy, and Relative Entropy (basic concepts of Statistical Mechanics) to the Induced Norm of Stochastic uncertain systems. Thus, establishing the precise connection between the principles of statistical mechanics and H^∞ or robustness of uncertain systems.
The second result of this paper is to relate the Free Energy of the system to Entropy Rate Functionals, via the Theory of Large Deviations. In particular, it is shown that the Cramer or Legendre-Fenchel Transform of Large Deviations is equivalent to the dual functional of a primal robustness problem.
The third result of the paper is to introduce max-plus measures using the entropy rate functional of Large Deviations, and then show that it is a measure of information, similar to that of Shannon (Average Entropy). Subsequently, to relate the Free Energy of a dynamical system to Storage functions arizing in dissipative systems.
The above results are applied to two robustness problems. The first problem is concerned with a class of uncertain stochastic control systems, in which the pay-off is described by the relative entropy between the nominal measure and the uncertain measure. The set of uncertain measures considered are described through energy inequality constraints expressed in terms of the uncertain measure. When stochastic dynamical systems are considered, this problem is equivalent to the sub-optimal disturbance attenuation problem arising in H^∞ control.[9–11]
With respect to this formulation, several properties of the optimal solution are discussed, and relations to Statistical Mechanics, Robust Control, Risk-Averse/Risk-Seeking strategies, and Cramer's Theorem of Large Deviations are identified. The implication of these results to minimax games and several monotonicity properties of the optimal strategies are derived.[9]
The second problem is concerned with the formulation of stochastic optimal control systems, in which uncertainty is described by a relative entropy constraint between the nominal measure and the uncertain measure, while the pay-off is a functional of the uncertain measure. This is a minimax game in which the controller seeks to minimize the pay-off, while the disturbance described by a set of measures aims at maximizing the pay-off. When stochastic dynamical systems are considered, this problem is equivalent to the optimal disturbance attenuation problem arising in H^∞ control.[17]
Some of the results presented in this paper have appeared in,[20, 21] while additional connections between the concepts under investigation can be found in.[?, 19]

2. STATISTICAL MECHANICS

Boltzman linked the microscopic properties of the particles that composed a thermodynamic system with their associate macroscopic properties by his celebrated equation

$$S = k \ln \Omega \tag{2.1}$$

where S is the entropy, macroscopic property, and Ω the number of microstates that are compatible with the state. Starting from this equation and using combinatorial arguments that are taking into consideration the various possible states that the system can access, in our case, n, it can be shown that the entropy is given by

$$S = -k \sum_i^n p_i^* \ln p_i^* \tag{2.2}$$

where $k = 1.3806503 \times 10^{-23} \frac{J}{K}$ is the Boltzmann constant and p_i^* is the probability of finding the system in state i.
For a thermodynamic system in contact with a heat bath of temperature T, Gibb's computed p_i^* to be,

$$p_i^* = \frac{e^{-E_i/kT}}{Q}; \qquad Q = \sum_{j=1}^n e^{-E_j/kT} \tag{2.3}$$

where Q is the partition function of the system. Substituting (2.3) in (2.2) the Helmhotz free energy F, a macroscopic thermodynamic quantity can be expressed in terms of the partition function Q as,

$$F = -kT \ln Q = -kT \log \sum_{j=1}^N e^{-E_j/kT} \tag{2.4}$$

Hence,

$$F = \sum_{i=1}^{n} E_i p_i^* + kT \sum_{i}^{n} p_i^* \ln p_i^* = U(p^*) - kTS(p^*) \tag{2.5}$$

Therefore, starting from the partition function, a mathematical structure that counts the number of possible microstates of a system the various thermodynamic quantities can be obtained.[20]

2.1. Variational Interpretation of Statistical Mechanics

The following theorem gives a variational interpretation of the basic principle of Statistical Mechanics. It will be used in subsequent sessions to related the statistical mechanics concepts to robustness concepts.

THEOREM 2.1. *Let Σ a non-empty denumerable set endowed with the discrete topology and $\mathcal{M}(\Sigma) = \Big\{\pi = (\pi_1, \ldots, \pi_N), \pi_j \geq 0, \sum_{j=1}^{N} \pi_j = 1, 1 \leq j \leq N\Big\}$.*
1) For every measurable function $E_j : \Sigma \to \Re, 1 \leq j \leq N$, and a fixed probability vector $\mu \in \mathcal{M}(\Sigma)$

$$\log\Big(\sum_{j=1}^{N} e^{\frac{-E_j(x)}{kT}} \mu_j(x)\Big)^{kT} = \sup_{\nu \in \mathcal{M}(\Sigma)} \Big\{ -\sum_{j=1}^{N} E_j(x)\nu_j(x) - kT \sum_{j=1}^{N} \frac{\nu_j(x)}{\mu_j(x)} \log \frac{\nu_j(x)}{\mu_j(x)} \Big\} \tag{2.6}$$

Moreover, the supremum is attained at

$$\nu_n^*(x) = \frac{e^{-E_n(x)/kT} \mu_n(x)}{\sum_{j=1}^{N} e^{-E_j(x)/kT} \mu_j(x)}, \qquad 1 \leq n \leq N \tag{2.7}$$

2) For every measurable function $E_j : \Sigma \to \Re, 1 \leq j \leq N$

$$\log\Big(\sum_{j=1}^{N} e^{\frac{-E_j(x)}{kT}}\Big)^{kT} = \sup_{\nu \in \mathcal{M}(\Sigma)} \Big\{ -\sum_{j=1}^{N} E_j(x)\nu_j(x) - kT \sum_{j=1}^{N} \nu_j(x) \log \nu_j(x) \Big\} \tag{2.8}$$

Moreover, the supremum is attained at

$$\nu_n^*(x) = \frac{e^{-E_n(x)/kT}}{\sum_{j=1}^{N} e^{-E_j(x)/kT}}, \qquad 1 \leq n \leq N \tag{2.9}$$

3) The basic principle of Statistical Mechanics, (2.8), (2.9) are the dual equations associated with the primal problem of maximizing the Entropy subject to an average energy constraint, defined by

$$\sup_{\nu \in \mathcal{M}(\Sigma)} \Big\{ -k \sum_{j=1}^{N} \nu_j(x) \log \nu_j(x) \Big\}; \qquad \textit{subject to} \quad \sum_{j=1}^{N} E_j(x)\nu_j(x) \leq \gamma, \gamma \in \Re \tag{2.10}$$

Proof. Follows from Theorem 3.1.

Therefore, the optimal measure ν^* can be characterised as the measure closest to the nominal measure (uniform measure in this case) and satisfies the constraint on the expected values of the random variables E_j.

3. ROBUSTNESS OF UNCERTAIN STOCHASTIC SYSTEMS

In this section the abstract formulation of the relevant problems together with a theorem that associates the tilted measure with this formulation are introduced.

3.1. Abstract Formulation

Let (Σ, d) denote a complete separable metric space, and $(\Sigma, \mathcal{B}(\Sigma))$ the corresponding measurable space in which $\mathcal{B}(\Sigma)$ are identified as the Borel sets generated by open sets in Σ. Let $\mathcal{M}(\Sigma)$ denote the set of probability measures on $(\Sigma, \mathcal{B}(\Sigma))$, $\mathcal{U}_{ad}$ the set of admissible controls, and $B(\Sigma;\Re)$ the set of bounded real-valued measurable functions, $\ell^u : \Sigma \rightarrow \Re$ for a given $u \in \mathcal{U}_{ad}$.

Here, $\mathcal{M}(\Sigma)$ denotes the set of all possible measures induced by the stochastic systems, while $\ell^u \in B(\Sigma;\Re)$ denotes the energy function or fidelity criterion associated with a given choice of the control law $u \in \mathcal{U}_{ad}$.

3.2. Tilted Measure

The following theorem characterises the tilted measure associated with the problems whose abstarct formulation is given above and it is employed in subsequent analysis.

THEOREM 3.1. *For every measurable function $\ell^u : \Sigma \rightarrow \Re$ bounded below and $\mu^u \in \mathcal{M}(\Sigma)$ and $s > 0$*

$$\log\Big(\int_\Sigma e^{\frac{\ell^u}{s}} d\mu^u\Big)^s = \sup_{\{\nu^u \in \mathcal{M}(\Sigma); H(\nu^u|\mu^u)<\infty\}} \Big\{\int_\Sigma \ell^u d\nu^u - sH(\nu^u|\mu^u)\Big\} \tag{3.11}$$

Moreover, if $\ell^u e^{\frac{\ell^u}{s}} \in L_1(\mu^u)$, then the supremum is attained at

$$d\nu^{u,*} = \frac{e^{\frac{\ell^u}{s}} d\mu^u}{\int_\Sigma e^{\frac{\ell^u}{s}} d\mu^u}$$

Proof. The derivation is similar to the one found in,[8] which treats the case $s = 1$.

It is clear that there is one to one relation between Theorem 2.1 and Theorem 3.1. In fact, Theorem 2.1 is a special case of Theorem 3.1, simply let $s \rightarrow kT, \ell^u \rightarrow -E, \mu^u \rightarrow \sum_{j=1}^N \delta(x-j)$. Therefore, any problem which is related to Theorem 3.1, it is also related to the Statistical Mechanics equations.

3.3. Robustness of Stochastic Uncertain Systems: An Energy Constraint Formulation

In this section, the first optimization problem described in the introduction is introduced.

3.3.1. Problem Statement

DEFINITION 3.2. *Let $u \in \mathcal{U}_{ad}$, let $\ell^u \in B(\Sigma; \mathcal{B}(\Sigma))$ (the space of bounded continuous functions), and $\mu^u \in \mathcal{M}(\Sigma)$ which is a fixed nominal measure, and $m \stackrel{\triangle}{=} E_{\mu^u} = \int_\Sigma \ell^u d\mu^u$, $\gamma \in \Re$.*
1) Find $\nu^{u,*} \in \mathcal{M}(\Sigma)$ which solves

$$J(u, \nu^{u,*}) = \inf_{\{\nu^u \in \mathcal{M}(\Sigma); \int_\Sigma \ell^u d\nu^u \leq \gamma, \quad H(\nu^u|\mu^u)<\infty\}} H(\nu^u|\mu^u) \tag{3.12}$$

for the following two cases.
Case 1. $m \stackrel{\triangle}{=} E_{\mu^u}(\ell^u) = \int_\Sigma \ell^u d\mu^u > \gamma$; Case 2. $m \stackrel{\triangle}{=} E_{\mu^u}(\ell^u) = \int_\Sigma \ell^u d\mu^u < \gamma$;
2) Find $\nu^{u,*} \in \mathcal{M}(\Sigma)$ which solves

$$J(u, \nu^{u,*}) = \inf_{\{\nu^u \in \mathcal{M}(\Sigma); \int_\Sigma \ell^u d\nu^u \geq \gamma, \quad H(\nu^u|\mu^u)<\infty\}} H(\nu^u|\mu^u) \tag{3.13}$$

for the following two cases.
Case 1. $m \stackrel{\triangle}{=} E_{\mu^u}(\ell^u) = \int_\Sigma \ell^u d\mu^u < \gamma$; Case 2. $m \stackrel{\triangle}{=} E_{\mu^u}(\ell^u) = \int_\Sigma \ell^u d\mu^u > \gamma$;

REMARK 3.3. *The fidelity constraints $E_{\nu^u}(\ell^u) \leq \gamma$, $E_{\nu^u}(\ell^u) \geq \gamma$ represent average energy constraints with respect to the unknown measure $\nu^u \in \mathcal{M}(\Sigma)$, such as integral quadratic constraints, tracking errors, etc., while*

γ is a parameter which is in some relation with $m \stackrel{\triangle}{=} E_{\mu^u}(\ell^u)$, that is, either $m > \gamma$ or $m < \gamma$. In particular, as shown in subsequent sections, the case (3.12), with $m > \gamma$ will correspond to the optimistic scenario (emphasizing the best cases) in which the strategies are risk-seeking, while the case (3.13), with $m < \gamma$ will correspond to the pessimistic scenario (emphasizing the worst cases) in which the strategies are risk-averse.

The above problems have various implications in minimax games, some of which are described below.

3.3.2. Related Problems

Disturbance Attenuation in Robustness. For a given $u \in \mathcal{U}_{ad}$ let $L_2(\nu^u;\mathcal{H}) \stackrel{\triangle}{=} \Big\{\phi^u : \Sigma \to \mathcal{H}; \phi^u$ is a measurable random variable such that $\int_\Sigma ||\phi||^2_{\mathcal{H}} d\nu^u < \infty\Big\}$ denote the Hilbert Space of random variables. Let $L_2(\nu^u;\mathcal{Z})$ and $L_2(\nu^u;\mathcal{D})$ denote the Hilbert Spaces of tracking signals and disturbance signals, respectively. For a given $u \in \mathcal{U}_{ad}$, let $T^u : \mathcal{D} \to \mathcal{Z}$ be a bounded linear operator with induced norm defined by

$$J(u) \stackrel{\triangle}{=} ||T^u|| = \sup_{||d||_{L_2(\nu^u;\mathcal{D})} \neq 0} \frac{||z||^2_{L_2(\nu^u;\mathcal{Z})}}{||d||^2_{L_2(\nu^u;\mathcal{D})}} \tag{3.14}$$

The sub-optimal disturbance attenuation is to ensure that for all $u \in \mathcal{U}_{ad}$ that $J(u) \leq \frac{1}{s}, s > 0$, which is equivalent to

$$J^s(u) = \sup_{d \in L_2(\nu^u;\mathcal{D})} \Big\{ s \int ||z||^2_{\mathcal{Z}} d\nu^u - \frac{1}{2}\int ||d||^2_{\mathcal{D}} d\nu^u \Big\} = - \inf_{d \in L_2(\nu^u;\mathcal{D})} \Big\{ \int ||d||^2_{\mathcal{D}} d\nu^u - s \int ||z||^2_{\mathcal{Z}} d\nu^u \Big\} \tag{3.15}$$

and ensuring that the pay-off is non-positive.
When ν^u is absolutely continuous with respect to μ^u, then it can be shown (see[11]) that $H(\nu^u|\mu^u) = \frac{1}{2}\int ||d||^2_{\mathcal{D}} d\nu^u$. Therefore, the dual functional associated with converting the primal problem (3.13) into the equivalent unconstrained optimization

$$J^{s,\gamma}(u, \nu^{u,*}) = \inf_{\nu^u \in \mathcal{M}(\Sigma)} \Big\{ H(\nu^u|\mu^u) - s\Big(E_{\nu^u}\Big(\ell^u\Big) - \gamma\Big)\Big\} \tag{3.16}$$

is equivalent to the sub-optimal disturbance attenuation problem (3.15) (let $\ell^u = ||z||^2_{\mathcal{Z}}$). Moreover, larger values of s imply higher attenuation and hence higher dissipation. An application of the above results to general nonlinear partially observable systems is discussed in.[11]

Legendre-Fenchel or Cramer Transform. In the context of large deviations, the dual functionals associated with converting the primal problems (3.12), (3.13) into equivalent unconstrained optimization problems are equal to the Legendre-Fenchel or Cramer transforms of ℓ^u defined by

$$I(\gamma) \stackrel{\triangle}{=} \sup_{s \in \Re} \Big\{ s\gamma - \log E_{\mu^u}\Big\{e^{s\ell^u}\Big\}\Big\} = \sup_{s \in \Re} \inf_{\nu^u \in \mathcal{M}(\Sigma)} \Big\{ H(\nu^u|\mu^u) - s\Big(E_{\nu^u}\Big(\ell^u\Big) - \gamma\Big)\Big\} \tag{3.17}$$

The Legendre-Fenchel or Cramer transform of ℓ^u is employed in Large Deviations Theory to identify the entropy rate functional $I(\gamma)$ associated with rare events.

Optimistic Versus Pessimistic Optimization. In the context of robust disturbance attenuation of uncertain systems, the measure $\mu^u \in \mathcal{M}(\Sigma)$ corresponds to the nominal measure, $\nu^u \in \mathcal{M}(\Sigma)$ corresponds to the uncertain measure, and the fidelity constraints $E_{\nu^u}\Big\{\ell^u\Big\} \leq \gamma$ and $E_{\nu^u}\Big\{\ell^u\Big\} \geq \gamma$ represent average energy constraints.
It can be shown that[9]:

1. for (3.12) the average energy constraint with respect to all uncertain measures $\nu^u << \mu^u$, $E_{\nu^u}\Big\{\ell^u\Big\} \leq \gamma$, is below the average energy of the nominal model, $E_{\mu^u}\Big\{\ell^u\Big\} > \gamma$; hence it represents an optimistic scenario.

2. for (3.13) the average energy constraint with respect to all uncertain measures $\nu^u << \mu^u$, $E_{\nu^u}\Big\{\ell^u\Big\} \geq \gamma$ is above the average energy of the nominal model $E_{\mu^u}\Big\{\ell^u\Big\} < \gamma$; hence it represents a pessimistic scenario.

The parameter $s \in \Re$ is the Lagrange multiplier associated with the dual functional of the primal problems (3.12), (3.13). In particular, $s \leq 0$ corresponds to (3.12) while $s \geq 0$ corresponds to (3.13).

Risk-Averse Versus Risk-Seeking Optimization. In the context of risk-sensitive pay-offs, (3.12) corresponds to an optimistic pay-off functional (emphasizing the best cases) in which the strategies are risk-seeking, and (3.13) corresponds to a pessimistic pay-off functional (emphasizing the worst cases) in which the strategies are risk-averse. Moreover, in the context of uncertain stochastic systems, risk-averse strategies always imply dissipation inequalities.

3.4. Robustness of Stochastic Uncertain Systems: A Relative Entropy Constraint Formulation

In this section the second optimization problem described in the introduction is considered.

3.4.1. Problem Statement

DEFINITION 3.4. *Let $u \in \mathcal{U}_{ad}$, ℓ^u measurable and bounded below, $\ell^u \in L_1(\mu^u)$, $\mu^u \in \mathcal{M}(\Sigma)$ which is a fixed nominal measure, and $R \in (0, \infty)$.*
Find $\nu^{u,*} \in \mathcal{M}(\Sigma)$ which achieves the supremum

$$J(u, \nu^{u,*}) = \sup_{\{\nu^u \in \mathcal{M}(\Sigma); H(\nu^u|\mu^u) \leq R\}} \int_\Sigma \ell^u d\nu^u, \qquad R \in (0, \infty) \tag{3.18}$$

Next, for every $s \in \Re$, define the Lagrangian associated with the problem of Definition 3.4

$$J^{s,R}(u, \nu^u) \stackrel{\triangle}{=} E_{\nu^u}(\ell^u) - s\Big(H(\nu^u|\mu^u) - R\Big) \tag{3.19}$$

and its associated dual functional

$$J^{s,R}(u, \nu^{u,*}) = \sup_{\{\nu^u \in \mathcal{M}(\Sigma)\}} J^s(u, \nu^u) \tag{3.20}$$

3.4.2. Related Problems

Disturbance Attenuation in Robustness. For a given $u \in \mathcal{U}_{ad}$, let $T^u : \mathcal{D} \to \mathcal{Z}$ be a bounded linear operator with induced norm defined by

$$J(u, d^*) \stackrel{\triangle}{=} ||T^u|| = \sup_{||d||_{L_2(\nu^u;\mathcal{D})} \neq 0} \frac{||z||^2_{L_2(\nu^u;\mathcal{Z})}}{||d||^2_{L_2(\nu^u;\mathcal{D})}} = \sup_{\frac{1}{2}||d||_{L_2(\nu^u;\mathcal{D})} \leq R} ||z||^2_{L_2(\nu^u;\mathcal{Z})} \tag{3.21}$$

Then the optimal control $u^* \in \mathcal{U}_{ad}$ is found by minimizing the induced norm

$$J(u^*, d^*) \stackrel{\triangle}{=} \inf_{u \in \mathcal{U}_{ad}} ||T^{u^*}|| = \inf_{u \in \mathcal{U}_{ad}} \sup_{\frac{1}{2}||d||_{L_2(\nu^u;\mathcal{D})} \leq R} ||z||^2_{L_2(\nu^u;\mathcal{Z})} \tag{3.22}$$

Moreover, the optimal control $u^* \in \mathcal{U}_{ad}$ is found by minimizing the induced norm, and it is given by

$$J^{s^*}(u^*, d^*) = \inf_{u \in \mathcal{U}_{ad}} \inf_{s \geq 0} \sup_{d \in L_2(\nu^u;\mathcal{D})} \Big\{ \int ||z||^2_{\mathcal{Z}} d\nu^u - s\Big(\frac{1}{2}\int ||d||^2_{\mathcal{D}} d\nu^u - R\Big)\Big\} \tag{3.23}$$

in which $\inf_{u\in\mathcal{U}_{ad}}\inf_{s\geq 0}$ is interchanged.
When ν^u is absolutely continuous with respect to μ^u, and the nominal model is described by stochastic differential equations which are driven by Brownian motion or general Martingales, then it can be shown that $H(\nu^u|\mu^u) = \frac{1}{2}\int ||d||^2_{\mathcal{D}} d\nu^u$. In this case, the primal problem (3.22) and its dual problem (3.23) are equivalent to the problem of Definition 3.4 (let $\ell^u = ||z||^2_{\mathcal{Z}}$). Moreover, smaller the values of s the higher the attenuation and hence the higher dissipation of otput power with respect to the input power.[19]

Risk-Averse Versus Risk-Seeking Optimization. In the context of risk-sensitive pay-offs, the problem of Definition 3.4 corresponds to an optimistic pay-off functional (emphasizing the best cases), when the Lagrange multiplier $s \leq 0$, in which the strategies are risk-seeking, and to a pessimistic pay-off functional (emphasizing the worst cases) in which the strategies are risk-averse, when the Lagrange multiplier $s \geq 0$.

4. LARGE DEVIATIONS THEORY AND RELATED DETERMINISTIC MEASURES

In this section, we start with the abstract formulation of the Large Deviations (LD) problems (see[22–24]). We seek to identify deterministic measures and axioms which emerge from the LD theory. In the process we will broach key results of LD theory to information theoretic deterministic measures, and in subsequent sections to the extremal theory of H^∞ optimization.

Throughout we let $\mathcal{X}$ be a Polish space (e.g., complete separable metric space), $\mathcal{B}_\mathcal{X}$ the Borel algebra of $\mathcal{X}$, and $\{P^\epsilon\}_{\epsilon>0}$ a family of probability measures on $\mathcal{B}_\mathcal{X}$. LD theory studies the deviant behaviors of the events $\mathcal{O} \in \mathcal{B}_\mathcal{X}$ for which $x_0 \notin \mathcal{O}$, in terms of the rate at which $P^\epsilon(\mathcal{O}) \to 0$, as $\epsilon \to 0$, by seeking expressions of the form

$$P^\epsilon(\mathcal{O})) = \exp\Big\{\frac{I(x)}{\epsilon} + o(1)\Big\}, \quad I : \mathcal{X} \to [-\infty, 0] \tag{4.24}$$

To illustrate the application of LD we consider a simple example in which $\{P^\epsilon\}_{\epsilon>0}$ are absolutely continuous with respect to some fixed measure Q, so we can write $dP^\epsilon(x) = C^\epsilon \exp\Big\{\frac{I(x)}{\epsilon}\Big\}dQ, \quad I : \mathcal{X} \to [-\infty, 0,$ where $I(x) = 0$ if and only if $x = x_0$ (x_0 being a typical element of $\mathcal{X}$). Assuming $\lim_{\epsilon\to 0}\epsilon \log C^\epsilon = 0$ uniformly, then for any $\mathcal{C} \in \mathcal{B}_\mathcal{X}$ with $Q(\mathcal{C}) < \infty$, we have

$$\epsilon \log P^\epsilon(\mathcal{C}) = \log\Big(\int_\mathcal{C} C^\epsilon \exp\Big(\frac{I(x)}{\epsilon}\Big)dQ\Big)^\epsilon; \qquad \lim_{\epsilon\to 0}\epsilon \log P^\epsilon(\mathcal{C}) = \log ess\sup\Big\{\exp\Big(I(x)\Big); x \in \mathcal{C}\Big\}$$

Hence, for a family $\{P^\epsilon\}_{\epsilon>0}$ which is absolutely continuous with respect to Q, denoted by $P^\epsilon << Q, \forall\epsilon > 0$, on $(\mathcal{X}, \mathcal{B}_\mathcal{X})$, which satisfy $Q(\mathcal{C}) < \infty$, for any $\mathcal{C} \in \mathcal{B}_\mathcal{X}$, define

$$\mu(\mathcal{C}) \stackrel{\triangle}{=} \lim_{\epsilon\to 0}\epsilon \log P^\epsilon(\mathcal{C}) = ess\sup\Big\{I(x); x \in \mathcal{C}\Big\} \tag{4.25}$$

Then for countable unions $\mathcal{C} = \bigcup_{i=1}^\infty \mathcal{A}_i$ we have

$$\mu(\mathcal{C}) = ess\sup\Big\{I(x); x \in \bigcup_{i=1}^\infty \mathcal{A}_i\Big\} = \sup_i\Big\{ess\sup\Big\{I(x); x \in \mathcal{A}_i\Big\}\Big\}, \quad \mathcal{A}_i \bigcap \mathcal{A}_j = \emptyset, \forall i \neq j, \quad \forall\{A_i\} \in \mathcal{B}_\mathcal{X}. \tag{4.26}$$

We note that if $(\mathcal{X}, \mathcal{B}_\mathcal{X}) = (\Re^d, \mathcal{B}(\Re^d))$ and $dQ(x) = d\ell_{\Re^d}(x)$ the Lebesgue measure on $\Re^d$ and

$$dP^\epsilon(x) = \frac{1}{(2\pi\epsilon)^{\frac{d}{2}}}\exp\Big(-\frac{||x||^2_{\Re^d}}{2\epsilon}\Big)d\ell_{\Re^d}(x) \tag{4.27}$$

then

$$\mu(\mathcal{C}) \stackrel{\triangle}{=} \lim_{\epsilon\to 0}\epsilon \log P^\epsilon(\mathcal{C}) = ess\sup\Big\{-\frac{||x||^2_{\Re^d}}{2} : x \in \mathcal{C}\Big\}, \quad \mathcal{C} \in \mathcal{B}_\mathcal{X} \tag{4.28}$$

For disjoint open sets $\{\mathcal{A}_i\} \in \mathcal{X}$ we have

$$\mu(\mathcal{C}) = \sup\Big\{I(x); x \in \bigcup_{i=1}^{n} \mathcal{A}_i\Big\} = \sup_i \Big\{\sup\Big\{I(x); x \in \mathcal{A}_i\Big\}\Big\}, \quad \mathcal{A}_i \bigcap \mathcal{A}_j = \emptyset, \forall i \neq j, \quad \forall\{A_i\} \in \mathcal{X}. \tag{4.29}$$

Therefore, the density associated with the measure defined by (4.28) is $I(x) = -\frac{||x||^2_{\Re^d}}{2}$. Guided by the above introduction, next we introduce the precise conditions for an underlying family of probability spaces to satisfy the LDP.

Definition 4.1. *(Large Deviations Principle).*[22–24] *Let* $\Big\{\Big(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, P^{\epsilon}\Big)\Big\}_{\epsilon>0}$ *be a family of complete probability spaces indexed by* ϵ *and let*

$$\mu^{\epsilon}_{\mathcal{X}}(\mathcal{A}) = \epsilon \log P^{\epsilon}(\mathcal{A}), \quad \mu_{\mathcal{X}}(\mathcal{A}) \stackrel{\triangle}{=} \lim_{\epsilon\to 0} \mu^{\epsilon}_{\mathcal{X}}(\mathcal{A}), \quad \mathcal{A} \in \mathcal{B}_{\mathcal{X}}$$

provided the limit exists.
We say that this probability space satisfies the Large Deviations Principle (LDP) with real-valued rate function $I_{\mathcal{X}}(\cdot)$, *denoted by* $\Big\{\Big(\mathcal{X}, \mathcal{B}_{\mathcal{X}}\ , P^{\epsilon}\Big)\Big\}_{\epsilon>0} \sim I_{\mathcal{X}}(x)$ *if there exists a function* $I_{\mathcal{X}} : \mathcal{X} \to [-\infty, 0]$ *called the action functional which satisfies the following properties.*

1. $-\infty \leq I_{\mathcal{X}}(x) \leq 0, \forall x \in \mathcal{X}$
2. $I_{\mathcal{X}}(\cdot)$ *is Upper Semicontinuous (u.s.c))*
3. *For each* $\mathcal{C} \in \mathcal{B}_{\mathcal{X}}$

$$\lim_{\epsilon\to 0} \sup \epsilon \log P^{\epsilon}(\mathcal{C}) \leq \sup_{x\in\bar{\mathcal{C}}} I_{\mathcal{X}}(x) \tag{4.30}$$

   where $\bar{\mathcal{C}}$ *is the closure of the set* $\mathcal{C} \in \mathcal{B}_{\mathcal{X}}$.

4. *For each* $\mathcal{C} \in \mathcal{B}_{\mathcal{X}}$

$$\lim_{\epsilon\to 0} \inf \epsilon \log P^{\epsilon}(\mathcal{C}) \geq \sup_{x\in\mathcal{O}^0} I_{\mathcal{X}}(x) \tag{4.31}$$

   where $\mathcal{C}^0$ *is the interior of the set* $\mathcal{C} \in \mathcal{B}_{\mathcal{X}}$.

5. *If* $\mathcal{C} \in \mathcal{B}_{\mathcal{X}}$ *is such that*

$$\sup_{x\in\mathcal{C}^0} I_{\mathcal{X}}(x) = \sup_{x\in\mathcal{C}} I_{\mathcal{X}}(x) = \sup_{x\in\bar{\mathcal{C}}} I_{\mathcal{X}}(x) \tag{4.32}$$

   then

$$\mu_{\mathcal{X}}(\mathcal{C}) = \sup_{x\in\mathcal{C}} I_{\mathcal{X}}(x) \tag{4.33}$$

   and $\mathcal{C} \in \mathcal{X}$ *is called a continuity set of* $I_{\mathcal{X}}(\cdot)$. *If (4.32) holds for all elements* $\mathcal{C} \in \mathcal{B}_{\mathcal{X}}$ *then* $\mathcal{B}_{\mathcal{X}}$ *is called a continuity* $\sigma-$*algebra of* $I_{\mathcal{X}}(\cdot)$.
   In 3, 4, 5 the supremum over an empty set is defined to be $-\infty$.

Note that $P^{\epsilon}(\mathcal{X}) = 1, \forall\epsilon > 0$ implies that $\sup_{x\in\mathcal{X}} I_{\mathcal{X}}(x) = 0$, and hence there exists at least one $x_0 \in \mathcal{X}$ for which $I_{\mathcal{X}}(x_0) = 0$. If $\mathcal{A} \in \mathcal{B}_{\mathcal{X}}$ is a set of measure zero, that is, $P^{\epsilon}(\mathcal{A}) = 0, \forall\epsilon > 0$, then $\sup_{x\in\mathcal{A}^0} I_{\mathcal{X}}(x) = -\infty$, which is the supremum of a functional over an empty set. Moreover, $P^{\epsilon}(\mathcal{A}) \in [0, 1], \forall\epsilon > 0, \mathcal{A} \in \mathcal{B}_{\mathcal{X}}$, implies that $-\infty \leq \sup_{x\in\mathcal{A}^0} I_{\mathcal{X}}(x) \leq \sup_{x\in\bar{\mathcal{A}}} I_{\mathcal{X}}(x) \leq 0$. Thus, for continuity $\sigma-$algebras, the measure $\mu_{\mathcal{X}}(\cdot)$ defined above has the properties of a (max,plus) measure.
Next, we shall introduce two fundamental Theorems associated with the LDP, which will lead to the conclusion that if $\mathcal{B}_{\mathcal{X}}$ is a family of continuity sets of $I_{\mathcal{X}}(\cdot)$ then $\mu_{\mathcal{X}}(\cdot)$ define above is a deterministic (max,plus) measure.

Assumptions 4.2. *In subsequent discussions, and unless otherwise state, we assume that all LD statements are with respect to continuity sets of* $I_{\mathcal{X}}(\cdot)$, *so that (4.32) is satisfied and (4.33) is well defined.*

4.1. Deterministic Measures

Armed with the above statements we shall show that if the family of probability spaces $\left\{\left(\mathcal{X},\mathcal{B}_\mathcal{X},P^\epsilon\right)\right\}_{\epsilon>0}$ satisfies the LDP with rate $I_\mathcal{X}(\cdot)$, then the rate induces a (max,plus) measure defined by

$$\mu_\mathcal{X}(\mathcal{O}) \stackrel{\triangle}{=} \lim_{\epsilon\to 0}\epsilon\log P^\epsilon(\mathcal{O}),\quad \forall \mathcal{O}\in\mathcal{B}_\mathcal{X} \tag{4.34}$$

Indeed, the countable additivity property of the measure for any finite collection $\{\mathcal{A}_i\}_{i=1}^n$ of disjoint sets in $\mathcal{B}_\mathcal{X}$, follows from the statements preceding Lemma 3.2, since

$$\begin{aligned}&\mu_\mathcal{X}\Big(\bigcup_{i=1}^n\mathcal{A}_i\Big) = \lim_{\epsilon\to 0}\epsilon\log\int_{\mathcal{A}_1\bigcup\ldots\bigcup\mathcal{A}_n}dP^\epsilon(x) = \lim_{\epsilon\to 0}\epsilon\log\Big(\sum_{i=1}^n\int_{\mathcal{A}_i}dP^\epsilon\Big)\\ &= \sup_{1\le i\le n}\Big\{I_\mathcal{X}(\mathcal{A}_i)\Big\},\quad I_\mathcal{X}(\mathcal{A}_i)\stackrel{\triangle}{=}\mu_\mathcal{X}(\mathcal{A}_i) = \sup\Big\{I_\mathcal{X}(x); x\in\mathcal{A}_i\Big\}\end{aligned} \tag{4.35}$$

Thus, $\mu_\mathcal{X}(\mathcal{A})\stackrel{\triangle}{=} I_\mathcal{X}(\mathcal{A})$ is a max-plus measure, e.g., it satisfies

$$i)\mu_\mathcal{X}(\mathcal{A})\in[-\infty,0],\quad \forall\mathcal{A}\in\mathcal{B}_\mathcal{X};\quad ii)\mu_\mathcal{X}(\Omega)=0;\quad iii)\mu_\mathcal{X}\Big(\bigcup_{i=1}^n\mathcal{A}_i\Big) = \bigoplus_{i=1}^n\mu_\mathcal{X}(\mathcal{A}_i),\quad \mathcal{A}_i\bigcap\mathcal{A}_j=\emptyset,\forall i\ne j,\quad \mathcal{A}_j\in\mathcal{B}_\mathcal{X}$$

Thus, each $x\in\mathcal{X}$ is associate with $I_\mathcal{X}(x)$, which is the self-rate functional. The more likely $x\in\mathcal{X}$ is the larger the value of $I_\mathcal{X}(x)\in[-\infty,0]$.

4.2. Information Theoretic Measures

In this section we illustrate the importance of the rate functionals in defining entropy which is another form of Shannon entropy.

4.2.1. Shannon Entropy

One of the fundamental concepts of information theory is the concept of entropy which is a measure of uncertainty of a R.V. Let $\left\{\left(\mathcal{X},\mathcal{B}_\mathcal{X},P_X^\epsilon\right)\right\}_{\epsilon>0}$ be a family of probability spaces, and $X:(\Omega,\mathcal{F})\to(\mathcal{X},B_\mathcal{X})$ a R.V. defined on it. Suppose that X is the output of a discrete information source having a finite alphabet containing M symbols, $\mathcal{X}=\Big\{x_1,x_2,\ldots,x_M\Big\}$, and each x_i is produced according to the probability $P_X^\epsilon(\{x_i\}), 1\le i\le M$. If x_i occurs then the amount of information associated with the known occurrence of x_i is defined by $-\log P_X^\epsilon(\{x_i\})$.
Hence, the average amount of information per source output symbol, known as the average information, uncertainty or entropy is

$$H(P_X^\epsilon) = -\sum_{i=1}^M P_X^\epsilon(\{x_1\})\log P_X^\epsilon(\{x_1\}),\quad \text{bits/symbol} \tag{4.36}$$

4.2.2. Entropy Rate Functional as a Measure of Information

Next, we shall introduce an entropy function defined with respect to the (max,plus) algebra and the rate functional associated with the LDP.

Definition 4.3. *Let* $\left\{\left(\mathcal{X}\times\mathcal{Y},\mathcal{B}_{\mathcal{X}\times\mathcal{Y}},P_{X,Y}^\epsilon\right)\right\}_{\epsilon>0}\sim I_{\mathcal{X},\mathcal{Y}}:\mathcal{X}\times\mathcal{Y}\to[-\infty,0]$.
The Entropy Rate Functional of any event $\mathcal{O}_x\in\mathcal{B}_\mathcal{X}$ *denoted by* $H_\mathcal{X}(\mathcal{O}_x)$ *is defined by*

$$H_\mathcal{X}(\mathcal{O}_x)\stackrel{\triangle}{=}\lim_{\epsilon\to 0}\epsilon\log\frac{1}{P_X^\epsilon\Big(X\in\mathcal{O}_x\Big)} = -\sup\Big\{I_\mathcal{X}(x);x\in\mathcal{O}_x\Big\} = -\mu_\mathcal{X}(\mathcal{O}_x),\ \mathcal{O}_x\in\mathcal{B}_\mathcal{X} \tag{4.37}$$

According to the above definition the entropy rate functional stated in (4.37) enjoys analogous properties as the entropy $H(P^\epsilon_X)$, defined by (4.36). Suppose $\mu_{\mathcal{X}}(A) = \sup\Big\{I_{\mathcal{X}}(x); x \in \mathcal{O}_x\Big\}$, $\mathcal{O}_x \in \mathcal{B}_{\mathcal{X}}$ is a (max,plus) measure induced by a variable X taking values in the discrete space, $\mathcal{X} = \Big\{x_1, x_2, \ldots, x_M\Big\}$, in which X models the output of a discrete information source, producing symbols according to the (max,plus) law $\Big\{I_{\mathcal{X}}(\{x_i\})\Big\}$. If symbos x_i occurs then the amount of information associated with the known occurrence of x_i is defined by $-I_{\mathcal{X}}(\{x_i\}) \geq 0$. Moreover, we have the following properies. i) The entropy rate functional is nonnegative, $H_{\mathcal{X}}(\mathcal{O}_x) \geq 0$, $\forall \mathcal{O}_x \in \mathcal{B}_{\mathcal{X}}$, and equal to zero, $H_{\mathcal{X}}(\mathcal{O}_x) = 0$, if and only if at least one $I_{\mathcal{X}}(\{x_i\})$, is equal to zero. Moreover, unlike the entropy function $H(P^\epsilon_X)$ which can be negative for continuous R.V., the entropy rate functional is never negative, because the rate functional $I_{\mathcal{X}} : \mathcal{X} \rightarrow [-\infty, 0]$. ii) The entropy rate functional $H_{\mathcal{X}}(\mathcal{O}_x)$ is a continuous function of the rate functional $I_{\mathcal{X}}$. iii) The entropy rate functional, $H_{\mathcal{X}}(\mathcal{O}_x)$, is a concave function of the rate finctional, $I_{\mathcal{X}}(x)$.

4.3. The LDP of Diffusion Processes

In this section we construct the action functional and a deterministic measure on cylinder sets of a Hilbert space. This is a consequence of the Large Deviation principle applied to Brownian motion found in.[22–24]

Assumptions 4.4. *$f : \Re^n \rightarrow \Re$, $\sigma : \Re^n \rightarrow \Re^n \otimes \Re^n$ are uniformly Lipschitz continuous, σ is bounded and $a(x) \stackrel{\triangle}{=} \sigma(x)\sigma'(x)$ is positive definite, that is, there exists an $k \in [1, \infty)$ such that*

$$||f(x) - f(y)|| + ||\sigma(x) - \sigma(y)|| \leq k||x - y||, \quad ||\sigma(x)|| \leq k, \quad \exists \lambda > 0 \ni \sigma(x)\sigma(x) \geq \lambda I_{n\times n}.$$

The LDP associated with diffusion processes is usually applied to the space $\Big(\mathcal{X}, \mathcal{B}_{\mathcal{X}}\Big) = \Big(C_{0,T} \stackrel{\triangle}{=} C([0,T]; \Re^n), \mathcal{B}_{0,T} \stackrel{\triangle}{=} \mathcal{B}(C([0,T]; \Re^n))\Big)$, which is a Banach space with the uniform norm $|| \cdot ||_{C_{0,T}}$. The diffusion process $\{X^\epsilon(t)\} : C_{0,T} \rightarrow C_{0,T}$ is the unique solution of the stochastic Ito differential equation

$$dX^\epsilon(t) = f(X^\epsilon(t))dt + \sqrt{\epsilon}\sigma(X^\epsilon(t))dw(t), \quad 0 \leq t \leq T, \quad X^\epsilon(0) = x \in \Re^n, \tag{4.38}$$

where validity of Assumption 4.4 is assumed. For a given bounded function f let $\{P^\epsilon\}_{\epsilon>0}$ denote the probability measure induced by $\{X^\epsilon(t)\}$ on $\Big(C_{0,T}, \mathcal{B}_{0,T}\Big)$. Then $P^\epsilon = \mathcal{W}^\epsilon \circ X^{\epsilon,-1}$ where $\mathcal{W}^\epsilon$ is the measure induced by $\{\sqrt{\epsilon}w(t)\}$ and $X^\epsilon : C_{0,T} \rightarrow C_{0,T}$ is defined by $X^\epsilon = F^\epsilon(g)$, where X^ϵ is the unique continuous solution of $X^\epsilon(t) = x + \int_0^t f(X^\epsilon(s))ds + g^\epsilon(t)$.
Introduce the Hilbert space

$$H^1_{0,T} = H^1\Big([0,T]; \Re^n\Big) \stackrel{\triangle}{=} \Big\{\phi \in C([0,T]; \Re^n); \phi(t) = \int_0^t \dot{\phi}(s)ds, \int_0^T ||\dot{\phi}(s)||^2_{\Re^n} ds < \infty\Big\} \tag{4.39}$$

which is the space of absolutely continuous functions with square-integrable derivatives.
Then $\Big\{\Big(C_{0,T}, \mathcal{B}_{0,T}, P^\epsilon\Big)\Big\}_{\epsilon>0}$ satisfies the LDP, which is an application of the contraction principle; the action functional is given by

$$I^{x,f}_{H^1_{0,T}}(X) = \left\{ \begin{array}{ll} -\frac{1}{2}\int_0^T ||a^{-\frac{1}{2}}(X(s))\Big(\dot{X}(s) - f(X(s))\Big)||^2_{\Re^n} ds, & X - x \in H^1_{0,T} \\ -\infty, & X - x \notin H^1_{0,T} \end{array} \right. \tag{4.40}$$

Equivalently,

$$I^{x,f}_{H^1_{0,T}}(X) = I^{x,0}_{H^1_{0,T}}(w) = \left\{ \begin{array}{ll} -\frac{1}{2}\int_0^T ||a^{-\frac{1}{2}}(X(s))\dot{w}(s)||^2_{\Re^n} ds, & w \in H^{1,w}_{0,T} \\ -\infty, & w \notin H^{1,w}_{0,T} \end{array} \right. \tag{4.41}$$

where $H^{1,w}_{0,T} \stackrel{\triangle}{=} \Big\{w \in H^1_{0,T}; X(t) = x + \int_0^t f(X(s))ds + \int_0^t \sigma(X(s))\dot{w}(s)\Big\}$.
Moreover, for any $\mathcal{A}_{0,t} \in \mathcal{B}_{0,t}$ which is a continuity set of $I^{x,f}_{H^1_{0,T}}(\cdot)$ we have

$$\mu_{\mathcal{X}}(\mathcal{A}_{0,t}) \stackrel{\triangle}{=} \lim_{\epsilon \rightarrow 0} \log P^\epsilon(\mathcal{A}_{0,t}) = \sup\Big\{I^{x,f}_{H^1_{0,t}}(X); X \in \mathcal{A}_{0,t}, X(0) = x\Big\}. \tag{4.42}$$

4.3.1. Connection to Thermodynamic Entropy

Next, for the class of systems described in Section 4.3, we employ Large Deviations to relate the Free Energy $\log \int_{C_{0,T}} e^{\int_0^T \ell(X)} dP^\epsilon(X)$ to the Macroscopic Thermodynamic Entropy S of Section 2.
Clearly, for any bounded and continuous function ℓ, by the Laplace-Varadhan Theorem of Large Deviations[22] we have

$$S(x) \stackrel{\triangle}{=} \lim_{\epsilon \to 0} \epsilon \log \int_{C_{0,T}} e^{\int_0^T \ell(X)} dP_x^\epsilon(X) = \sup_{w \in H_{0,T}^{1,w}} \left\{ \int_0^T \left(\ell(X(s)) - \frac{1}{2} ||a^{-\frac{1}{2}}(X(s)) \dot{w}(s)||_{\Re^n}^2 \right) ds \right\} \tag{4.43}$$

Here $W_R(x, \dot{w}) \stackrel{\triangle}{=} \frac{1}{2} ||a^{-\frac{1}{2}}(x)\dot{w}||_{\Re^n}^2 - \ell(x)$ is the supply of energy into the system, and $S(x)$ is a storage function.[2]

5. CONCLUSION

This paper establishes various connections between Robustness, Information Theory, Large Daviations and Statistical Mechanics for stochastic uncertain systems. Detail derivations and additional properties and connections among these fields are found in.[11, 19, 20]

Acknowledgements. The authors are grateful to Professor Costas Georghiades of Texas A & M University for extending the invitation to contribute the material of this paper.

REFERENCES

1. G. Zames, "Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses," *IEEE Transactions on Automatic Control*, vol. 26, no. 2, pp. 301–320, 1981.
2. J. Willems, "Dissipative dynamical systems part I: General theory," *Arch. Rational Mech. Anal.*, vol. 45, pp. 321–351, 1972.
3. D. Jacobson, "Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games," *IEEE Transactions on Automatic Control*, vol. 18, no. 2, pp. 124–131, 1973.
4. A. Bensoussan and J. H. van Schuppen, "Optimal control of partially observable stochastic systems with an exponential-of-integral performance index," *SIAM Journal on Control and Optimization*, vol. 23, no. 4, pp. 599–613, 1985.
5. P. Whittle, "A risk-sensitive maximum principle: The case of imperfect state observations," *IEEE Transactions on Automatic Control*, vol. 36, no. 7, pp. 793–801, 1991.
6. C. Charalambous and J. Hibey, "Minimum principle for partially observable nonlinear risk-sensitive control problems using measure-valued decompositions," *Stochastics and Stochastics Reports*, vol. 57, 1996.
7. C. Charalambous, "The role of informations state and adjoint in relating nonlinear output feedback risk-sensitive control and dynamic games," *IEEE Transactions on Automatic Control*, vol. 42, no. 8, pp. 1163–1170, 1997.
8. P. D. Pra, L. Meneghini, and W. Runggaldier, "Some connections between stochastic control and dynamic games." *Mathematics of Control, Signals, and Systems* , vol. 9, pp. 303–326, 1996.
9. C. D. Charalambous and F. Rezaei, "Optimization of Stochastic Uncertain Systems: Large Deviations and Robustness." *in Proceedings of the 42nd IEEE Conference on Decision and Control*, Hawaii, U.S.A., December 9-12, Vol.4, pages 4249-4253, 2003.
10. F. Rezaei and C. D. Charalambous and A. Kyprianou, "Optimization of Fully Observable Nonlinear Stochastic Uncertain Controlled Systems." *in Proceedings of the 43nd IEEE Conference on Decision and Control,* Bahamas, December 14-17, 2004.
11. C. D. Charalambous and F. Rezaei and N. Ahmed, "Optimization of Stochastic Uncertain Systems: Entropy Rate Functionals, Cramer's Theorem, Minimax Games and Robustness." *SIAM Journal on Control and Optimization, (submitted),* pages-40, April 2004.

12. A. Isidori and A. Astolfi, "Disturbance attenuation and $\mathcal{H}_\infty$−control via measurement feedback in nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 37, no. 9, pp. 1283–1293, 1992.
13. P. Whittle, "A risk-sensitive maximum principle," *Systems and Control Letters*, vol. 15, pp. 183–192, 1990.
14. W. H. Fleming and W. M. McEneaney, "Risk-sensitive control and differential games, pp.185-197," in *Stochastic Theory and Adaptive Control, eds., T.E. Duncan and B. Pasik-Duncan*, (New York), Springer-Verlag 1992.
15. M. James, J. Baras, and R. Elliott, "Risk-sensitive control and dynamic games for partially observed discrete-time nonlinear systems," *IEEE Transaction on Automatic Control*, vol. 39, no. 4, pp. 780–792, 1994.
16. I. Petersen and M. R. James and P. Dupuis "Minimax optimal control of stochastic uncertain systems with relative entropy constrains," *IEEE Transactions on Automatic Control*, vol. 45, pp. 398–412, 2000.
17. C. D. Charalambous and F. Rezaei, "Characterization of the Optimal Disturbance Attenuation for Nonlinear Stochastic Uncertain Systems," *in Proceedings of the 42nd IEEE Conference on Decision and Control*, Hawaii, U.S.A., December 9-12, Vol.4, pages 4260-4264, 2003.
18. F. Rezaei and C. D. Charalambous and A. Kyprianou, "Optimization of Fully Observable Nonlinear Stochastic Uncertain Controlled Systems," *in Proceedings of the 43nd IEEE Conference on Decision and Control*, Bahamas, December 14-17, 2004.
19. C. D. Charalambous and F. Rezaei, "Stochastic Uncertain Systems Subject to Relative Entropy Constraints: Induced Norms and Monotonicity Properties of Minimax Games," *IEEE Transactions on Automatic Control (submitted)*, pages-38, December 29, 2004.
20. C. D. Charalambous and F. Rezaei and A. Kyprianou, "Relations Between Information Theory, Robustness and Statistical Mechanics of Stochastic Systems," *in Proceedings of the 43nd IEEE Conference on Decision and Control*, Bahamas, December 14-17, 2004.
21. W.M. McEneaney and C.D. Charalambous, "Large Deviations Theory, Induced Log-Plus and Max-Plus Measures and their Applications," *in Mathematical Theory of Network and Systems 2000,* Perpignan, France, June 19-23, 2000.
22. J.-D. Deuschel and D.W. Stroock, *Large Deviations,* Academic Press Inc., 1989.
23. S.R.S. Varadhan, *Large Deviations and Applications,* Society for Industrial and Applied Mathematics, 1984.
24. D.W. Stroock, *An Introduction to the Theory of Large Deviations,* Springer-Verlag, 1984.