Extremum Problems With Total Variation Distance and Their Applications

Charalambos D. Charalambous, Ioannis Tzortzis, Sergey Loyka, and Themistoklis Charalambous

Abstract—**The aim of this paper is to investigate extremum problems with pay-off being the total variation distance metric defined on the space of probability measures, subject to linear functional constraints on the space of probability measures, and vice-versa; that is, with the roles of total variation metric and linear functional interchanged. Utilizing concepts from signed measures, the extremum probability measures of such problems are obtained in closed form, by identifying the partition of the support set and the mass of these extremum measures on the partition. The results are derived for abstract spaces; specifically, complete separable metric spaces known as Polish spaces, while the high level ideas are also discussed for denumerable spaces endowed with the discrete topology. These extremum problems often arise in many areas, such as, approximating a family of probability distributions by a given probability distribution, maximizing or minimizing entropy subject to total variation distance metric constraints, quantifying uncertainty of probability distributions by total variation distance metric, stochastic minimax control, and in many problems of information, decision theory, and minimax theory.**

Index Terms—**Extremum probability measures, signed measures, total variation distance.**

I. Introduction

TOTAL variation distance metric on the space of probability measures is a fundamental quantity in statistics and probability, which over the years appeared in many diverse applications. In information theory it is used to define strong typicality and asymptotic equipartition of sequences generated by sampling from a given distribution [1]. In decision problems, it arises naturally when discriminating the results of observation of two statistical hypotheses [1]. In studying the ergodicity of Markov Chains, it is used to define the Dobrushin coefficient and establish the contraction property of transition probability distributions [2]. Moreover, distance in total variation of probability measures is related via upper and lower bounds to an anthology of distances and distance metrics [3]. The measure of distance in total variation of probability measures is a strong form of closeness of probability measures, and, convergence with respect to total variation of probability measures implies their convergence with respect to other distances and distance metrics.

In this paper, we formulate and solve several extremum problems involving the total variation distance metric and we discuss their applications in the areas of control, communication and statistics. The main problems investigated are:

(a) Extremum problems of linear functionals on the space of measures subject to a total variation distance metric constraint defined on the space of measures.
(b) Extremum problems of total variation distance metric on the space of measures subject to a linear functional constraint on the space of measures.
(c) Applications of these extremum problems, and their relations to other problems.

The formulation of these extremum problems, their discussion in terms of applications, and the contributions of this paper are developed at the abstract level, in which systems are represented by probability distributions on abstract spaces (complete separable metric space, known as Polish spaces [4]), pay-offs are represented by linear functionals on the space of probability measures or by distance in variation of probability measures, and constraints by linear functionals or distance in variation of probability measures. We consider Polish spaces since they are general enough to handle various models of practical interest, such as stochastic control problems on Borel spaces.

Utilizing concepts from signed measures, closed form expressions of the probability measures are derived which achieve the extremum of these problems. The construction of the extremum measures involves the identification of the partition of their support set, and their mass defined on these partitions. Throughout the derivations we make extensive use of lower and upper bounds of pay-offs which are achievable. Several simulations are carried out to illustrate the different features of the extremum solution of the various problems. An interesting observation concerning one of the extremum problems is its equivalent formulation as an extremum problem involving the oscillator semi-norm of the pay-off functional. The formulation and results obtained for these problems at the abstract level are discussed throughout the paper in the context of various applications, including the following.

(i) Dynamic Programming Under Uncertainty, to deal with uncertainty of transition probability distributions, via minimax theory, with total variation distance metric uncertainty constraints to codify the impact of incorrect distribution models on performance of the optimal strategies [5]. This formulation is applicable to Markov decision

Manuscript received January 14, 2013; revised November 14, 2013 and March 28, 2014; accepted May 1, 2014. Date of publication May 5, 2014; date of current version August 20, 2014. Recommended by Associate Editor L. H. Lee.

C. D. Charalambous and I. Tzortzis are with the Department of Electrical Engineering, University of Cyprus, Nicosia, Cyprus (e-mail: chadcha@ucy.ac.cy; tzortzis.ioannis@ucy.ac.cy).

S. Loyka is with the School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON, Canada (e-mail: sergey.loyka@ieee.org).

T. Charalambous is with the School of Electrical Engineering, Royal Institute of Technology (KTH), Stockholm, Sweden (e-mail: themisc@kth.se).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2014.2321951

problems subject to uncertainty. A specific example is presented in Section VII.

(ii) Approximation of Probability Distributions with Total Variation Distance Metric, to approximate a given high dimensional probability distribution μ on a measurable space $(\Sigma, \mathcal{B}(\Sigma))$ by another lower dimensional distribution ν on $(\overline{\Sigma}, \mathcal{B}(\overline{\Sigma}))$, $\overline{\Sigma} \subseteq \Sigma$, via minimization of the total variation distance metric between them subject to linear functional constraints. Model and graph reduction can be handled via such approximations. Graphs, for example, constitute the foundation of many real-world datasets. However, the size of the graph can become prohibitive to understand essential information that they contain. The reduction of graph-based models is significant in a wide variety of applications, such as placement of autonomous sensors, modeling Central Processing Unit (CPU) and database demands in web-based software engineering, and identifying the evolution in clusters within massive dynamic datasets in database research.

(iii) Maximization or Minimization of Entropy Subject to Total Variation Distance Metric Constraints, to invoke insufficient reasoning based on maximizing the entropy $H(\nu)$ of an unknown probability distribution ν on denumerable space Σ subject to a constraint on the total variation distance metric.

The rest of the paper is organized as follows. In Section II, total variation distance is defined, the extremum problems are introduced, while several related problems are discussed together with their applications. In Section III, some of the properties of the problems are discussed. In Sections III-A and III-B, signed measures are utilized to convert the extremum problems into equivalent ones, and to characterize the extremum measures on abstract spaces. In Section IV, closed form expressions of the extremum measures are derived for finite alphabet spaces. In Section V, the relation between total variation distance and other distance metrics is discussed. In Section VI, several simulations are presented to illustrate how the optimal distribution of the extremum problems behaves, for different scenarios of the support set of the distribution. Finally, in Section VII, an example from the area of stochastic optimal control is presented to demonstrate how the results obtained can be applied in practice.

II. Extremum Problems

In this section, we will introduce the extremum problems we investigate in this paper. Let (Σ, d_Σ) denote a complete, separable metric space and $(\Sigma, \mathcal{B}(\Sigma))$ the corresponding measurable space, where $\mathcal{B}(\Sigma)$ is the σ-algebra generated by open sets in Σ. Let $\mathcal{M}_1(\Sigma)$ denote the set of probability measures on $\mathcal{B}(\Sigma)$. The total variation distance[1] is a metric [6] $d_{TV} : \mathcal{M}_1(\Sigma) \times \mathcal{M}_1(\Sigma) \to [0, \infty)$ defined by

$$d_{TV}(\alpha, \beta) \equiv \|\alpha - \beta\|_{TV} \triangleq \sup_{P \in \mathcal{P}(\Sigma)} \sum_{F_i \in P} |\alpha(F_i) - \beta(F_i)| \quad (1)$$

where $\alpha, \beta \in \mathcal{M}_1(\Sigma)$ and $\mathcal{P}(\Sigma)$ denotes the collection of all finite partitions of Σ. With respect to this metric, $(\mathcal{M}_1(\Sigma), d_{TV})$ is a complete metric space. Since the elements of $\mathcal{M}_1(\Sigma)$ are probability measures, then $d_{TV}(\alpha, \beta) \leq 2$. In minimax problems one can introduce an uncertainty set based on distance in variation as follows. Suppose the probability measure $\nu \in \mathcal{M}_1(\Sigma)$ is unknown, while modeling techniques give access to a nominal probability measure $\mu \in \mathcal{M}_1(\Sigma)$. Having constructed the nominal probability measure, one may construct from empirical data, the distance of the two measures with respect to the total variation distance $\|\nu - \mu\|_{TV}$. This will provide an estimate of the radius R, such that $\|\nu - \mu\|_{TV} \leq R$, and hence characterize the set of all possible true measures $\nu \in \mathcal{M}_1(\Sigma)$, centered at the nominal distribution $\mu \in \mathcal{M}_1(\Sigma)$, and lying within the ball of radius R, with respect to the total variation distance $\|\cdot\|_{TV}$. Such a procedure is used in information theory to define strong typicality of sequences. Unlike other distances used in the past such as relative entropy [7]–[11], quantifying uncertainty via the metric $\|\cdot\|_{TV}$ does not require absolute continuity of measures,[2] i.e., singular measures are admissible, and hence ν and μ need not be defined on the same space. Thus, the support set of μ may be $\tilde{\Sigma} \subset \Sigma$, hence $\mu(\Sigma \setminus \tilde{\Sigma}) = 0$ but $\nu(\Sigma \setminus \tilde{\Sigma}) \neq 0$ is allowed. For measures induced by stochastic differential equations (SDE's), variation distance uncertainty set models situations in which both the drift and diffusion coefficient of SDE's are unknown.

Define the spaces $BC(\Sigma) \triangleq \{$Bounded continuous functions $\ell : \Sigma \mapsto \mathbb{R} : \|\ell\| \triangleq \sup_{x \in \Sigma} |\ell(x)| < \infty\}$, $BC^+(\Sigma) \triangleq \{\ell \in BC(\Sigma) : \ell \geq 0\}$, $BM(\Sigma) \triangleq \{$Bounded measurable functions $\ell : \Sigma \mapsto \mathbb{R} : \|\ell\| < \infty\}$, $BM^+(\Sigma) \triangleq \{\ell \in BM(\Sigma) : \ell \geq 0\}$. $BC(\Sigma)$ and $BM(\Sigma)$ endowed with the sup norm $\|\ell\| \triangleq \sup_{x \in \Sigma} |\ell(x)|$, are Banach spaces [6]. Next, we introduce the two main extremum problems we shall investigate in this paper.

Problem 2.1: Given a fixed nominal distribution $\mu \in \mathcal{M}_1(\Sigma)$ and a parameter $R \in [0, 2]$, define the class of true distributions by

$$\mathbb{B}_R(\mu) \triangleq \{\nu \in \mathcal{M}_1(\Sigma) : \|\nu - \mu\|_{TV} \leq R\} \quad (2)$$

and the average pay-off with respect to the true probability measure $\nu \in \mathbb{B}_R(\mu)$ by

$$\mathbb{L}_1(\nu) \triangleq \int_\Sigma \ell(x)\nu(dx), \quad \ell \in BC^+(\Sigma) \text{ or } BM^+(\Sigma). \quad (3)$$

The objective is to find the solution of the extremum problem

$$D^+(R) \triangleq \sup_{\nu \in \mathbb{B}_R(\mu)} \int_\Sigma \ell(x)\nu(dx). \quad (4)$$

Problem 2.1 is a convex optimization problem on the space of probability measures. Note that, $BC^+(\Sigma)$, $BM^+(\Sigma)$ can be generalized to $L^{\infty,+}(\Sigma, \mathcal{B}(\Sigma), \nu)$, the set of all $\mathcal{B}(\Sigma)$-measurable, non-negative essentially bounded functions

[1] The definition of total variation distance can be extended to signed measures.

[2] $\nu \in \mathcal{M}_1(\Sigma)$ is absolutely continuous with respect to $\mu \in \mathcal{M}_1(\Sigma)$, denoted by $\nu \ll \mu$, if $\mu(A) = 0$ for some $A \in \mathcal{B}(\Sigma)$ then $\nu(A) = 0$.

defined $\nu - a.e.$ endowed with the essential supremum norm $\|\ell\|_{\infty,\nu} = \nu\text{-ess} \quad \sup_{x\in\Sigma} \ell(x) \triangleq \inf_{\Delta\in\mathcal{N}_\nu} \sup_{x\in\Delta^c} \|\ell(x)\|$, where $\mathcal{N}_\nu = \{A \in \mathcal{B}(\Sigma) : \nu(A) = 0\}$.

In the context of minimax theory, Problem 2.1 is important in minimax stochastic control, estimation, and decision. Such formulations are found in [7]–[11] utilizing relative entropy to describe a class of models, and in [12], [13] utilizing L_1 distance to describe a class of power spectral densities. An application of Problem 2.1 in the context of dynamic programming is presented in Section VII.

The second extremum problem is defined below.

Problem 2.2: Given a fixed nominal distribution $\mu \in \mathcal{M}_1(\Sigma)$ and a parameter $D \in [0, \infty)$, define the class of true distributions by

$$\mathbb{Q}(D) \triangleq \left\{ \nu \in \mathcal{M}_1(\Sigma) : \int_\Sigma \ell(x)\nu(dx) \leq D \right\} \tag{5}$$

where $\ell \in BC^+(\Sigma)$ or $BM^+(\Sigma)$, and the total variation pay-off with respect to the true probability measure $\nu \in \mathbb{Q}(D)$ by

$$\mathbb{L}_2(\nu) \triangleq \|\nu - \mu\|_{TV}. \tag{6}$$

The objective is to find the solution of the extremum problem

$$R^-(D) \triangleq \inf_{\nu\in\mathbb{Q}(D)} \|\nu - \mu\|_{TV} \tag{7}$$

whenever[3] $\int_\Sigma \ell(x)\mu(dx) > D$.

Problem 2.2 is important in the context of approximation theory, since distance in variation is a measure of proximity of two probability distributions subject to constraints. It is also important in spectral measure or density approximation as follows. Recall that a function $\{R(\tau) : -\infty \leq \tau \leq \infty\}$ is the covariance function of a quadratic mean continuous and wide-sense stationary process if and only if it is of the form [14]

$$R(\tau) = \int_{-\infty}^{\infty} e^{2\pi\nu\tau} F(d\nu),$$

where $F(\cdot)$ is a finite Borel measure on $\mathbb{R}$, called spectral measure. Thus, by proper normalization of $F(\cdot)$ via $F_N(d\nu) \triangleq (1/R(0))F(d\nu)$, then $F_N(d\nu)$ is a probability measure on $\mathcal{B}(\mathbb{R})$, and hence Problem 2.2 can be used to approximate the class of spectral measures with moment estimates belonging to the class described by inequality constraints. Spectral estimation problems are discussed extensively in [15]–[19], utilizing relative entropy and Hellinger distances, under moment estimates involving equality constraints. However, in these references, the approximated spectral density is absolutely continuous with respect to the nominal spectral density; hence, it can not deal with reduced order approximation. In this respect, distance in total variation between spectral measures is very attractive.

[3]If $\int_\Sigma \ell(x)\mu(dx) \leq D$ then $\nu^* = \mu$ is the trivial extremum measure of (7).

A. Related Extremum Problems

Problems 2.1, 2.2 are related to additional extremum problems which are introduced below.

1) The solution of (4) gives the solution to the problem defined by

$$R^+(D) \triangleq \sup_{\nu\in\mathcal{M}_1(\Sigma): \int_\Sigma \ell(x)\nu(dx)\leq D} \|\nu - \mu\|_{TV}. \tag{8}$$

Specifically, $R^+(D)$ is the inverse mapping of $D^+(R)$. $D^+(R)$ is investigated in [20] in the context of minimax stochastic control, following an alternative approach which utilizes large deviation theory to express the extremum measure by a convex combination of a tilted and the nominal probability measures. The two disadvantages of the method pursued in [8]–[11] are the following. 1) No explicit closed form expression for the extremum measure is given, and as a consequence, 2) its application to dynamic programming is restricted to a class of uncertain probability measures which are absolutely continuous with respect to the nominal measure $\mu(\Sigma) \in \mathcal{M}_1(\Sigma)$.

2) Let ν and μ be absolutely continuous with respect to the Lebesgue measure so that $\varphi(x) \triangleq (d\nu/d(x))(x)$, $\psi(x) \triangleq (d\mu/dx)(x)$ (e.g., $\varphi(\cdot)$, $\psi(\cdot)$ are the probability density functions of $\nu(\cdot)$ and $\mu(\cdot)$, respectively. Then, $\|\nu - \mu\|_{TV} = \int_\Sigma |\varphi(x) - \psi(x)|dx$ and hence, (4) and (8) are L_1-distance optimization problems.

3) Let Σ be a non-empty denumerable set endowed with the discrete topology including finite cardinality $|\Sigma|$, with $\mathcal{M}_1(\Sigma)$ identified with the standard probability simplex in $\mathbb{R}^{|\Sigma|}$, that is, the set of all $|\Sigma|$-dimensional vectors which are probability vectors, and $\ell(x) \triangleq -\log\nu(x)$, $x \in \Sigma$ (emerges from an additional constraint—Kraft inequality), where $\{\nu(x) : x \in \Sigma\} \in \mathcal{M}_1(\Sigma)$, $\{\mu(x) : x \in \Sigma\} \in \mathcal{M}_1(\Sigma)$. Then, (4) is equivalent to maximizing the entropy of $\{\nu(x) : x \in \Sigma\}$ subject to total variation distance metric constraint defined by

$$D^+(R) = \sup_{\nu\in\mathcal{M}_1(\Sigma): \sum_{x\in\Sigma}|\nu(x)-\mu(x)|\leq R} H(\nu). \tag{9}$$

Problem (9) is of interest when the concept of insufficient reasoning (e.g., Jayne's maximum entropy principle [21], [22]) is applied to construct a model for $\nu \in \mathcal{M}_1(\Sigma)$, subject to information quantified via total variation distance metric between ν and an empirical distribution μ. In the context of stochastic control systems for a class of distributions, and its relation to robustness, Problem (9) with the total variation distance constraint replaced by relative entropy distance constraint is investigated in [23], [24].

4) The solution of (7) gives the solution to the problem defined by

$$D^-(R) \triangleq \inf_{\nu\in\mathcal{M}_1(\Sigma): \|\nu-\mu\|_{TV}\leq R} \int_\Sigma \ell(x)\nu(dx). \tag{10}$$

Problems (7) and (10) are important in approximating a class of probability distributions or spectral measures by reduced ones. In fact, the solution of (10) is obtained precisely as that of Problem 2.1, with a reverse computation of the partition of the space Σ and the mass of the extremum measure on the partition moving in the opposite direction.

III. Characterization of Extremum Measures on Abstract Spaces

This section utilizes signed measures and some of their properties to convert Problems 2.1, 2.2 into equivalent extremum problems. We describe the results using abstract spaces to avoid excluding measures defined on Borel spaces. First, we discuss some of the properties of these extremum Problems.

Lemma 3.1:

1) $D^+(R)$ is a non-decreasing concave function of R, and

$$D^+(R) = \sup_{\|\nu-\mu\|_{TV}=R} \int_\Sigma \ell(x)\nu(dx), \quad \text{if } R \le R_{\max} \tag{11}$$

where $R_{\max}$ is the smallest non-negative number belonging to [0, 2] such that $D^+(R)$ is constant in $[R_{\max}, 2]$.

2) $R^-(D)$ is a non-increasing convex function of D, and

$$R^-(D) = \inf_{\int_\Sigma \ell(x)\nu(dx)=D} \|\nu-\mu\|_{TV}, \quad \text{if } D \le D_{\max} \tag{12}$$

where $D_{\max}$ is the smallest non-negative number belonging to $[0,\infty)$ such that $R^-(D)=0$ for any $D \in [D_{\max},\infty)$.

Proof:

1) Suppose $0 \le R_1 \le R_2$, then for every $\nu \in \mathbb{B}_{R_1}(\mu)$ we have $\|\nu-\mu\|_{TV} \le R_1 \le R_2$, and therefore $\nu \in \mathbb{B}_{R_2}(\mu)$, hence

$$\sup_{\nu\in\mathbb{B}_{R_1}(\mu)} \int_\Sigma \ell(x)\nu(dx) \le \sup_{\nu\in\mathbb{B}_{R_2}(\mu)} \int_\Sigma \ell(x)\nu(dx)$$

which is equivalent to $D^+(R_1) \le D^+(R_2)$. So $D^+(R)$ is a non-decreasing function of R. Now consider two points $(R_1, D^+(R_1))$ and $(R_2, D^+(R_2))$ on the linear functional curve, such that $\nu_1 \in \mathbb{B}_{R_1}(\mu)$ achieves the supremum of (4) for R_1, and $\nu_2 \in \mathbb{B}_{R_2}(\mu)$ achieves the supremum of (4) for R_2. Then, $\|\nu_1-\mu\|_{TV} \le R_1$ and $\|\nu_2-\mu\|_{TV} \le R_2$. For any $\lambda \in (0,1)$, we have

$$\|\lambda\nu_1 + (1-\lambda)\nu_2 - \mu\|_{TV} \le \lambda\|\nu_1-\mu\|_{TV} + (1-\lambda)\|\nu_2-\mu\|_{TV} \le \lambda R_1 + (1-\lambda)R_2 = R.$$

Define $\nu^* \triangleq \lambda\nu_1 + (1-\lambda)\nu_2$, $R \triangleq \lambda R_1 + (1-\lambda)R_2$. The previous equation implies that $\nu^* \in \mathbb{B}_R(\mu)$, hence $D^+(\lambda R_1 + (1-\lambda)R_2) \ge \int_\Sigma \ell(x)\nu^*(dx)$. Therefore

$$\begin{aligned} D^+(R) &\ge \int_\Sigma \ell(x)\nu^*(dx) \\ &= \int_\Sigma \ell(x)\left(\lambda\nu_1(dx) + (1-\lambda)\nu_2(dx)\right) \\ &= \lambda\int_\Sigma \ell(x)\nu_1(dx) + (1-\lambda)\int_\Sigma \ell(x)\nu_2(dx) \\ &= \lambda D^+(R_1) + (1-\lambda)D^+(R_2). \end{aligned}$$

So, $D^+(R)$ is a concave function of R. Also the right side of (11), say $\bar{D}^+(R)$, is concave function of R. But $D^+(R) = \sup_{R'\le R} \bar{D}^+(R')$ which completes the derivation of (11).

2) The derivation is similar to (1). ■

Let $\mathcal{M}_{sm}(\Sigma)$ denote the set of finite signed measures. Then, any $\eta \in \mathcal{M}_{sm}(\Sigma)$ has a Jordan decomposition [25] $\{\eta^+, \eta^-\}$ such that $\eta = \eta^+ - \eta^-$, and the total variation of η is defined by $\|\eta\|_{TV} \triangleq \eta^+(\Sigma) + \eta^-(\Sigma)$. Define the following subset $\mathbb{M}_0(\Sigma) \triangleq \{\eta \in \mathcal{M}_{sm}(\Sigma) : \eta(\Sigma) = 0\}$. For $\xi \in \mathbb{M}_0(\Sigma)$, then $\xi(\Sigma)=0$, which implies that $\xi^+(\Sigma) = \xi^-(\Sigma)$, and hence $\xi^+(\Sigma) = \xi^-(\Sigma) = \|\xi\|_{TV}/2$. Then, $\xi \triangleq \nu - \mu \in \mathbb{M}_0(\Sigma)$ and hence $\xi = (\nu-\mu)^+ - (\nu-\mu)^- \equiv \xi^+ - \xi^-$.

A. Equivalent Extremum Problem of $D^+(R)$

Consider the pay-off of Problem 2.1, for $\ell \in BC^+(\Sigma)$. The solution is based on finding an upper bound which is achievable. Then the following inequalities hold.

$$\begin{aligned} \mathbb{L}_1(\nu) &\triangleq \int_\Sigma \ell(x)\nu(dx) \\ &\overset{(a)}{=} \int_\Sigma \ell(x)\left(\xi^+(dx) - \xi^-(dx)\right) + \int_\Sigma \ell(x)\mu(dx) \\ &\overset{(b)}{\le} \sup_{x\in\Sigma} \ell(x)\xi^+(\Sigma) - \inf_{x\in\Sigma} \ell(x)\xi^-(\Sigma) + \int_\Sigma \ell(x)\mu(dx) \\ &\overset{(c)}{=} \sup_{x\in\Sigma} \ell(x)\frac{\|\xi\|_{TV}}{2} - \inf_{x\in\Sigma} \ell(x)\frac{\|\xi\|_{TV}}{2} + \int_\Sigma \ell(x)\mu(dx) \\ &= \left\{\sup_{x\in\Sigma} \ell(x) - \inf_{x\in\Sigma} \ell(x)\right\}\frac{\|\xi\|_{TV}}{2} + \int_\Sigma \ell(x)\mu(dx) \end{aligned} \tag{13}$$

where (a) follows by adding and subtracting $\int_\Sigma \ell(x)\mu(dx)$, and from the Jordan decomposition of $(\nu-\mu)$, (b) follows due to $\ell \in BC^+(\Sigma)$, (c) follows because any $\xi \in \mathbb{M}_0(\Sigma)$ satisfies

$\xi^+(\Sigma) = \xi^-(\Sigma) = (1/2)\|\xi\|_{TV}$. For a given $\mu \in \mathcal{M}_1(\Sigma)$ and $\nu \in \mathbb{B}_R(\mu)$ define the set

$$\widetilde{\mathbb{B}}_R(\mu) \triangleq \{\xi \in \mathbb{M}_0(\Sigma) : \xi = \nu - \mu, \nu \in \mathcal{M}_1(\Sigma), \|\xi\|_{TV} \leq R\}.$$

The upper bound in the right hand side of (13) is achieved by $\xi^* \in \widetilde{\mathbb{B}}_R(\mu)$ as follows. Let

$$x^0 \in \Sigma^0 \triangleq \{x \in \Sigma : \ell(x) = \sup\{\ell(x) : x \in \Sigma\} \equiv M\}$$

$$x_0 \in \Sigma_0 \triangleq \{x \in \Sigma : \ell(x) = \inf\{\ell(x) : x \in \Sigma\} \equiv m\}.$$

Take

$$\xi^*(dx) = \nu^*(dx) - \mu(dx) = \frac{R}{2}\left(\delta_{x^0}(dx) - \delta_{x_0}(dx)\right) \quad (14)$$

where $\delta_y(dx)$ denotes the Dirac measure concentrated at $y \in \Sigma$. This is indeed a signed measure with total variation $\|\xi^*\|_{TV} = \|\nu^* - \mu\|_{TV} = R$, and $\int_\Sigma \ell(x)(\nu^* - \mu)(dx) = (R/2)(M - m)$. Hence, by using (14) as a candidate of the maximizing distribution then the extremum Problem 2.1 is equivalent to

$$D^+(R) = \left\{\sup_{x\in\Sigma} \ell(x) - \inf_{x\in\Sigma} \ell(x)\right\}\frac{R}{2} + \int_\Sigma \ell(x)\mu(dx) \quad (15)$$

where ν^* satisfies the constraint $\|\xi^*\|_{TV} = \|\nu^* - \mu\|_{TV} = R$, it is normalized $\nu^*(\Sigma) = 1$, and $0 \leq \nu^*(A) \leq 1$ on any $A \in \mathcal{B}(\Sigma)$. Alternatively, the pay-off $\int_\Sigma \ell(x)\nu^*(dx)$ can be written as

$$\int_\Sigma \ell(x)\nu^*(dx) = \int_{\Sigma^0} M\nu^*(dx) + \int_{\Sigma_0} m\nu^*(dx) + \int_{\Sigma\setminus\Sigma^0\cup\Sigma_0} \ell(x)\mu(dx). \quad (16)$$

Hence, the optimal distribution $\nu^* \in \mathbb{B}_R(\mu)$ satisfies

$$\int_{\Sigma^0} \nu^*(dx) = \mu(\Sigma^0) + \frac{R}{2} \in [0,1] \quad (17)$$

$$\int_{\Sigma_0} \nu^*(dx) = \mu(\Sigma_0) - \frac{R}{2} \in [0,1] \quad (18)$$

$$\nu^*(A) = \mu(A),\ \forall A \subseteq \Sigma \setminus \Sigma^0 \cup \Sigma_0. \quad (19)$$

Remark 3.2:

1) For $\mu \in \mathcal{M}_1(\Sigma)$ which do not include point mass, and for $f \in BC^+(\Sigma)$, if Σ^0 and Σ_0 are countable, then (19) is $\mu(\Sigma^0) = \mu(\Sigma_0) = 0$, $\nu^*(\Sigma_0) = 0$, $\nu^*(\Sigma^0) = R/2$, $\nu^*(\Sigma \setminus \Sigma^0 \cup \Sigma_0) = \mu(\Sigma \setminus \Sigma^0 \cup \Sigma_0) - (R/2) \in [0,1]$.
2) The first right side term in (15) is related to the oscillator seminorm of $f \in BM(\Sigma)$ called global modulus of continuity, defined by $\text{osc}(f) \triangleq \sup_{(x,y)\in\Sigma\times\Sigma} |f(x) - f(y)| = 2\inf_{\alpha\in\mathbb{R}} \|f - \alpha\|$. For $f \in BM^+(\Sigma)$, $\text{osc}(f) = \sup_{x\in\Sigma} |f(x)| - \inf_{x\in\Sigma} |f(x)|$.

B. Equivalent Extremum Problem of $R^-(D)$

Next, we proceed with the abstract formulation of Problem 2.2. Consider the constraint of Problem 2.2, for $\ell \in BC^+(\Sigma)$. Then the following inequalities hold:

$$\begin{aligned}
&\int_\Sigma \ell(x)\nu(dx) \\
&= \int_\Sigma \ell(x)\left(\xi^+(dx) - \xi^-(dx)\right) + \int_\Sigma \ell(x)\mu(dx) \\
&\geq \inf_{x\in\Sigma} \ell(x)\xi^+(\Sigma) - \sup_{x\in\Sigma} \ell(x)\xi^-(\Sigma) + \int_\Sigma \ell(x)\mu(dx) \\
&= \inf_{x\in\Sigma} \ell(x)\frac{\|\xi\|_{TV}}{2} - \sup_{x\in\Sigma} \ell(x)\frac{\|\xi\|_{TV}}{2} + \int_\Sigma \ell(x)\mu(dx) \\
&= \left\{\inf_{x\in\Sigma} \ell(x) - \sup_{x\in\Sigma} \ell(x)\right\}\frac{\|\xi\|_{TV}}{2} + \int_\Sigma \ell(x)\mu(dx). \quad (20)
\end{aligned}$$

The lower bound on the right hand side of (20) is achieved by choosing $\xi^* \in \widetilde{\mathbb{B}}_R(\mu)$ as follows:

$$\xi^*(dx) = \nu^*(dx) - \mu(dx) = \frac{R}{2}\left(\delta_{x_0}(dx) - \delta_{x^0}(dx)\right). \quad (21)$$

This is a signed measure with total variation $\|\xi^*\|_{TV} = \|\nu^* - \mu\|_{TV} = R$. Hence, by using (21) as a candidate of the minimizing distribution then (20) is equivalent to

$$\int_\Sigma \ell(x)\nu^*(dx) = \frac{R}{2}\left\{\inf_{x\in\Sigma} \ell(x) - \sup_{x\in\Sigma} \ell(x)\right\} + \int_\Sigma \ell(x)\mu(dx). \quad (22)$$

Solving the above equation with respect to R the extremum Problem 2.2 (for $D < \int_\Sigma \ell(x)\mu(dx)$) is equivalent to

$$R^-(D) = \frac{2\left(D - \int_\Sigma \ell(x)\mu(dx)\right)}{\left\{\inf_{x\in\Sigma} \ell(x) - \sup_{x\in\Sigma} \ell(x)\right\}} \quad (23)$$

where ν^* satisfies the constraint $\int_\Sigma \ell(x)\nu^*(dx) = D$, it is normalized $\nu^*(\Sigma) = 1$, and $0 \leq \nu(A) \leq 1$ on any $A \in \mathcal{B}(\Sigma)$. We can now identify $R_{\max}$ and $D_{\max}$ described in Lemma 3.1. These are stated as a corollary.

Corollary 3.3: The values of $R_{\max}$ and $D_{\max}$ described in Lemma 3.1 are given by

$$R_{\max} = 2\left(1 - \mu(\Sigma^0)\right) \text{ and } D_{\max} = \int_\Sigma \ell(x)\mu(dx).$$

Proof: Concerning $R_{\max}$, $D^+(R) \leq \sup_{x\in\Sigma} \ell(x)$, $\forall R \geq 0$, hence $D^+(R_{\max})$ can be at most $\sup_{x\in\Sigma} \ell(x)$. Since $D^+(R)$ is non-decreasing then $D^+(R_{\max}) \leq D^+(R) \leq \sup_{x\in\Sigma} \ell(x)$, for any $R \geq R_{\max}$. Consider a ν that achieves this supremum. Let $\mu(\Sigma^0)$ and $\nu(\Sigma^0)$ to denote the nominal and true probability measures on Σ^0, respectively. If $\nu(\Sigma^0)=1$ then $\nu(\Sigma\setminus\Sigma^0)=0$. Therefore

$$\begin{aligned}\|\nu-\mu\|_{TV} &= \sum_{x\in\Sigma^0}|\nu(x)-\mu(x)| + \sum_{x\in\Sigma\setminus\Sigma^0}|\nu(x)-\mu(x)| \\ &\overset{(a)}{=} \sum_{x\in\Sigma^0}\nu(x) - \sum_{x\in\Sigma^0}\mu(x) + \sum_{x\in\Sigma\setminus\Sigma^0}\mu(x) \\ &= 2\left(1-\sum_{x\in\Sigma^0}\mu(x)\right) = 2\left(1-\mu(\Sigma^0)\right) \qquad (24)\end{aligned}$$

where (a) follows due to $\nu(\Sigma\setminus\Sigma^0)=0$ which implies $\nu(x)=0$ for any $x\in\Sigma\setminus\Sigma^0$, and because $\nu(x)\ge\mu(x)$ for all $x\in\Sigma^0$. Therefore, $R_{\max}=2(1-\mu(\Sigma^0))$ implies that $D^+(R_{\max})=\sup_{x\in\Sigma}\ell(x)$. Hence, $D^+(R)=\sup_{x\in\Sigma}\ell(x)$, for any $R\ge R_{\max}$.

Concerning $D_{\max}$, $R^-(D)\ge 0$ for all $D\ge 0$ hence $R^-(D_{\max})$ can be at least zero. Let $D_{\max}=\int_\Sigma \ell(x)\mu(dx)$, then it is obvious that $R^-(D_{\max})=0$. Since $R^-(D)$ in non-increasing, then $0\le R^-(D)\le R^-(D_{\max})$, for any $D\ge D_{\max}$. Hence, $R^-(D)=0$, for any $D\ge D_{\max}$. ∎

IV. Characterization of Extremum Measures for Finite Alphabets

This section uses the results of Section III to compute closed form expressions for the extremum measures ν^* for any $R\in[0,2]$, when Σ is a finite alphabet space to give the intuition into the solution procedure. This is done by identifying the sets $\Sigma^0,\Sigma_0,\Sigma\setminus\Sigma^0\cup\Sigma_0$, and the measure ν^* on these sets for any $R\in[0,2]$. Although this can be done for probability measures on complete separable metric spaces (Polish spaces) (Σ,d_Σ), and for $\ell\in BM^+(\Sigma)$, $\ell\in BC^+(\Sigma)$, $L^{\infty,+}(\Sigma,\mathcal{B}(\Sigma),\nu)$, we prefer to discuss the finite alphabet case to gain additional insight into these problems. At the end of this section we shall use the finite alphabet case to discuss the extensions to countable alphabet and to $\ell\in L^{\infty,+}(\Sigma,\mathcal{B}(\Sigma),\nu)$.

Consider the finite alphabet case $(\Sigma,\mathcal{M})$, where $\mathrm{card}(\Sigma)=|\Sigma|$ is finite, $\mathcal{M}=2^{|\Sigma|}$. Thus, ν and μ are point mass distributions on Σ. Define the set of probability vectors on Σ by

$$\mathbb{P}(\Sigma)\triangleq\left\{p=(p_1,\ldots,p_{|\Sigma|}):p_i\ge 0, i=0,\ldots,|\Sigma|,\sum_{i\in\Sigma}p_i=1\right\}.$$

Thus, $p\in\mathbb{P}(\Sigma)$ is a probability vector in $\mathbb{R}_+^{|\Sigma|}$. Also let $\ell\triangleq\{\ell_1,\ldots,\ell_{|\Sigma|}\}$ so that $\ell\in\mathbb{R}_+^{|\Sigma|}$ (e.g., set of non-negative vectors of dimension $|\Sigma|$).

A. Problem 2.1: Finite Alphabet Case

Suppose $\nu\in\mathbb{P}(\Sigma)$ is the true probability vector and $\mu\in\mathbb{P}(\Sigma)$ is the nominal fixed probability vector. The extremum problem is defined by

$$D^+(R)\triangleq\max_{\nu\in\mathbb{B}_R(\mu)}\sum_{i\in\Sigma}\ell_i\nu_i \qquad (25)$$

where

$$\mathbb{B}_R(\mu)\triangleq\left\{\nu\in\mathbb{P}(\Sigma):\|\nu-\mu\|_{TV}\triangleq\sum_{i\in\Sigma}|\nu_i-\mu_i|\le R\right\}. \qquad (26)$$

Next, we apply the results of Section III to characterize the optimal ν^* for any $R\in[0,2]$. By defining $\xi_i\triangleq\nu_i-\mu_i$, $i=1,\ldots,|\Sigma|$, and $\xi\in\mathbb{M}_0(\Sigma)$, Problem 2.1 can be reformulated as follows:

$$\max_{\nu\in\mathbb{B}_R(\mu)}\sum_{i\in\Sigma}\ell_i\nu_i\longrightarrow\sum_{i\in\Sigma}\ell_i\mu_i+\max_{\xi\in\widetilde{\mathbb{B}}_R(\mu)}\sum_{i\in\Sigma}\ell_i\xi_i. \qquad (27)$$

Note that $\xi\in\widetilde{\mathbb{B}}_R(\mu)$ is described by the constraints

$$\alpha\triangleq\sum_{i\in\Sigma}|\xi_i|\le R,\quad \sum_{i\in\Sigma}\xi_i=0,\quad 0\le\xi_i+\mu_i\le 1,\ \forall i\in\Sigma. \qquad (28)$$

The positive and negative variation of the signed measure ξ are defined by $\xi^+=\max\{\xi,0\}$ and $\xi^-=\max\{-\xi,0\}$. Therefore

$$\sum_{i\in\Sigma}\xi_i=\sum_{i\in\Sigma}\xi_i^+-\sum_{i\in\Sigma}\xi_i^-,\quad \sum_{i\in\Sigma}|\xi_i|=\sum_{i\in\Sigma}\xi_i^++\sum_{i\in\Sigma}\xi_i^- \qquad (29)$$

and hence $\sum_{i\in\Sigma}\xi_i^+\equiv\alpha/2\equiv\sum_{i\in\Sigma}\xi_i^-$. In addition

$$\sum_{i\in\Sigma}\ell_i\xi_i=\sum_{i\in\Sigma}\ell_i\xi_i^+-\sum_{i\in\Sigma}\ell_i\xi_i^-. \qquad (30)$$

Define the maximum and minimum values of the sequence $\{\ell_1,\ldots,\ell_{|\Sigma|}\}$ by $\ell_{\max}\triangleq\max_{i\in\Sigma}\ell_i$, $\ell_{\min}\triangleq\min_{i\in\Sigma}\ell_i$, and its corresponding support sets by $\Sigma^0\triangleq\{i\in\Sigma:\ell_i=\ell_{\max}\}$, $\Sigma_0\triangleq\{i\in\Sigma:\ell_i=\ell_{\min}\}$. For all remaining sequence, $\{\ell_i:i\in\Sigma\setminus\Sigma^0\cup\Sigma_0\}$, and for $1\le r\le|\Sigma\setminus\Sigma^0\cup\Sigma_0|$ define recursively the set of indices for which ℓ achieves its $(k+1)$th smallest value by

$$\Sigma_k\triangleq\left\{i\in\Sigma:\ell_i=\min\left\{\ell_\alpha:\alpha\in\Sigma\setminus\Sigma^0\cup\left(\bigcup_{j=1}^{k}\Sigma_{j-1}\right)\right\}\right\} \qquad (31)$$

where $k\in\{1,2,\ldots,r\}$, till all the elements of Σ are exhausted (i.e., k is at most $|\Sigma\setminus\Sigma^0\cup\Sigma_0|$). Define the corresponding values of the sequence of sets in (31) by

$$\ell(\Sigma_k)\triangleq\min_{i\in\Sigma\setminus\Sigma^0\cup\left(\bigcup_{j=1}^{k}\Sigma_{j-1}\right)}\ell_i,\quad k\in\{1,2,\ldots,r\}$$

where r is the number of Σ_k sets which is at most $|\Sigma\setminus\Sigma^0\cup\Sigma_0|$; for example, when $k=1$, $\ell(\Sigma_1)=\min_{i\in\Sigma\setminus\Sigma^0\cup\Sigma_0}\ell_i$, when $k=2$, $\ell(\Sigma_2)=\min_{i\in\Sigma\setminus\Sigma^0\cup\Sigma_0\cup\Sigma_1}\ell_i$, etc. Note that if $\ell_1<\ell_2<\ldots<\ell_{|\Sigma|}$ then $\Sigma^0=\{|\Sigma|\}$, $\Sigma_0=\{1\}$ and $\Sigma_k=\{k+1\}$ for $k=1,\ldots,|\Sigma|-2$. The following theorem characterizes the solution of Problem 2.1.

Theorem 4.1: The solution of the finite alphabet version of Problem 2.1 is given by

$$D^+(R)=\ell_{\max}\nu^*(\Sigma^0)+\ell_{\min}\nu^*(\Sigma_0)+\sum_{k=1}^{r}\ell(\Sigma_k)\nu^*(\Sigma_k). \qquad (32)$$

Moreover, the optimal probabilities are given by

$$\nu^*(\Sigma^0) \triangleq \sum_{i\in\Sigma^0} \nu_i^* = \sum_{i\in\Sigma^0} \mu_i + \frac{\alpha}{2} \tag{33a}$$

$$\nu^*(\Sigma_0) \triangleq \sum_{i\in\Sigma_0} \nu_i^* = \left(\sum_{i\in\Sigma_0} \mu_i - \frac{\alpha}{2}\right)^+ \tag{33b}$$

$$\nu^*(\Sigma_k) \triangleq \sum_{i\in\Sigma_k} \nu_i^* = \left(\sum_{i\in\Sigma_k} \mu_i - \left(\frac{\alpha}{2} - \sum_{j=1}^{k} \sum_{i\in\Sigma_{j-1}} \mu_i\right)^+\right)^+ \tag{33c}$$

$$\alpha = \min(R, R_{\max}), \qquad R_{\max} \triangleq 2\left(1 - \sum_{i\in\Sigma^0} \mu_i\right) \tag{33d}$$

where, $k = 1, 2, \ldots, r$ and r is the number of Σ_k sets which is at most $|\Sigma \setminus \Sigma^0 \cup \Sigma_0|$.

The solution of Problem 2.1 is obtained by identifying the partition of Σ into disjoint sets $\{\Sigma^0, \Sigma_0, \Sigma_1, \ldots, \Sigma_k\}$ and the measures on this partition. The main idea is to express the total variation distance constraint as a summation of the positive and negative variation of a signed measure, and then to find upper and lower bounds on the probabilities of Σ^0 and $\Sigma \setminus \Sigma^0$, which are achievable. Utilizing the fact that the positive and negative variation parts of the total variation distance have equal mass concentrated on them, closed form expressions of the probability measures, on these sets, which achieve the upper and lower bounds are derived.

In the following Lemma upper and lower bounds which are achievable are obtained. These they will be used for the derivation of Theorem 4.1.

Lemma 4.2:

(a) Upper Bound.

$$\sum_{i\in\Sigma} \ell_i \xi_i^+ \le \ell_{\max}\left(\frac{\alpha}{2}\right). \tag{34}$$

The bound holds with equality if

$$\sum_{i\in\Sigma^0} \mu_i + \frac{\alpha}{2} \le 1, \qquad \sum_{i\in\Sigma^0} \xi_i^+ = \frac{\alpha}{2}, \qquad \xi_i^+ = 0,\ \forall i \in \Sigma \setminus \Sigma^0. \tag{35}$$

(b) Lower Bound.

Case 1) If $\sum_{i\in\Sigma_0} \mu_i - (\alpha/2) \ge 0$ then

$$\sum_{i\in\Sigma} \ell_i \xi_i^- \ge \ell_{\min}\left(\frac{\alpha}{2}\right). \tag{36}$$

The bound holds with equality if

$$\sum_{i\in\Sigma_0} \mu_i - \frac{\alpha}{2} \ge 0, \qquad \sum_{i\in\Sigma_0} \xi_i^- = \frac{\alpha}{2}, \qquad \xi_i^- = 0,\ \forall i \in \Sigma \setminus \Sigma_0. \tag{37}$$

Case 2) If $\sum_{j=1}^{k} \sum_{i\in\Sigma_{j-1}} \mu_i - (\alpha/2) \le 0$ for any $k \in \{1, 2, \ldots, r\}$ then

$$\sum_{i\in\Sigma} \ell_i \xi_i^- \ge \ell(\Sigma_k)\left(\frac{\alpha}{2} - \sum_{j=1}^{k} \sum_{i\in\Sigma_{j-1}} \mu_i\right) + \sum_{j=1}^{k} \sum_{i\in\Sigma_{j-1}} \ell_i \mu_i. \tag{38}$$

Moreover, equality holds if

$$\sum_{i\in\Sigma_{j-1}} \xi_i^- = \sum_{i\in\Sigma_{j-1}} \mu_i, \text{ for all } j = 1, 2, \ldots, k \tag{39a}$$

$$\sum_{i\in\Sigma_k} \xi_i^- = \left(\frac{\alpha}{2} - \sum_{j=1}^{k} \sum_{i\in\Sigma_{j-1}} \mu_i\right) \tag{39b}$$

$$\sum_{j=0}^{k} \sum_{i\in\Sigma_j} \mu_i - \frac{\alpha}{2} \ge 0 \tag{39c}$$

$$\xi_i^- = 0 \text{ for all } i \in \Sigma \setminus \Sigma_0 \cup \Sigma_1 \cup \ldots \cup \Sigma_k. \tag{39d}$$

Proof: Part (a) and Part (b), case 1, follows from Section III-A. For Part (b), case 2, we proceed as follows. Consider any $k \in \{1, 2, \ldots, r\}$. First, we show that inequality holds. From Part (b), case 1, we have that

$$\sum_{i\in\Sigma\setminus\cup_{j=1}^{k}\Sigma_{j-1}} \ell_i \xi_i^- \ge \min_{i\in\Sigma\setminus\cup_{j=1}^{k}\Sigma_{j-1}} \ell_i \sum_{i\in\Sigma\setminus\cup_{j=1}^{k}\Sigma_{j-1}} \xi_i^-$$
$$= \ell(\Sigma_k) \sum_{i\in\Sigma\setminus\cup_{j=1}^{k}\Sigma_{j-1}} \xi_i^- = \ell(\Sigma_k)\left(\sum_{i\in\Sigma} \xi_i^- - \sum_{j=1}^{k} \sum_{i\in\Sigma_{j-1}} \xi_i^-\right).$$

Hence,

$$\sum_{i\in\Sigma} \ell_i \xi_i^- - \sum_{j=1}^{k} \sum_{i\in\Sigma_{j-1}} \ell_i \xi_i^- \ge \ell(\Sigma_k)\left(\frac{\alpha}{2} - \sum_{j=1}^{k} \sum_{i\in\Sigma_{j-1}} \mu_i\right)$$

which implies

$$\sum_{i\in\Sigma} \ell_i \xi_i^- \ge \ell(\Sigma_k)\left(\frac{\alpha}{2} - \sum_{j=1}^{k} \sum_{i\in\Sigma_{j-1}} \mu_i\right) + \sum_{j=1}^{k} \sum_{i\in\Sigma_{j-1}} \ell_i \mu_i.$$

Next, we show under the stated conditions that equality holds.

$$\sum_{i\in\Sigma} \ell_i \xi_i^- = \sum_{j=1}^{k} \sum_{i\in\Sigma_{j-1}} \ell_i \xi_i^- + \sum_{i\in\Sigma_k} \ell_i \xi_i^- + \sum_{i\in\Sigma\setminus\cup_{j=0}^{k}\Sigma_j} \ell_i \xi_i^-$$
$$= \sum_{j=1}^{k} \ell(\Sigma_{j-1}) \sum_{i\in\Sigma_{j-1}} \xi_i^- + \ell(\Sigma_k) \sum_{i\in\Sigma_k} \xi_i^-$$
$$= \sum_{j=1}^{k} \sum_{i\in\Sigma_{j-1}} \ell_i \mu_i + \ell(\Sigma_k)\left(\frac{\alpha}{2} - \sum_{j=1}^{k} \sum_{i\in\Sigma_{j-1}} \mu_i\right).$$

■

Proof of Theorem 4.1: From Lemma 3.1, part (1), and Corollary 3.3, we know that for $R \le R_{\max}$, where $R_{\max} = 2(1 - \mu(\Sigma^0))$, the total variation constraint holds with equality, that is, $\|\xi\|_{TV} = R$. Let $\alpha = \|\xi\|_{TV}$. From (27) and (28), Problem 2.1 is given by

$$D^+(R) = \sum_{i\in\Sigma} \ell_i \mu_i + \max_{\xi\in\widetilde{\mathbb{B}}_R(\mu)} \sum_{i\in\Sigma} \ell_i \xi_i. \tag{40}$$

where $\xi \in \widetilde{\mathbb{B}}_R(\mu)$ is described by the constraints

$$\alpha \triangleq \sum_{i\in\Sigma} |\xi_i| = R, \quad \sum_{i\in\Sigma} \xi_i = 0, \quad 0 \le \xi_i + \mu_i \le 1, \quad \forall i \in \Sigma. \tag{41}$$

To maximize (40) we employ (30). It is obvious that an upper and a lower bound must be obtained for $\sum_{i\in\Sigma}\ell_i\xi_i^+$ and $\sum_{i\in\Sigma}\ell_i\xi_i^-$, respectively.

From Lemma 4.2, Part (a), the upper bound (34), holds with equality if conditions given by (35) are satisfied. Note that, $\sum_{i\in\Sigma^0}\mu_i+(\alpha/2)\leq 1$ is always satisfied and from the second equation of (35) we have that $\sum_{i\in\Sigma^0}\nu_i=\sum_{i\in\Sigma^0}\mu_i+(\alpha/2)$ and hence the optimal probability on Σ^0 is given by

$$\nu^*(\Sigma^0)\triangleq\sum_{i\in\Sigma^0}\nu_i^*=\sum_{i\in\Sigma^0}\mu_i+\frac{\alpha}{2}. \tag{42}$$

From Lemma 4.2, Part (b), case 1, the lower bound (36), holds with equality if conditions given by (37) are satisfied. Furthermore, from the second equation of (37) we have that $\sum_{i\in\Sigma_0}\nu_i=\sum_{i\in\Sigma_0}\mu_i-(\alpha/2)$ and condition given by the first equation of (37) must be satisfied, hence the optimal probability on Σ_0 is given by

$$\nu^*(\Sigma_0)\triangleq\sum_{i\in\Sigma_0}\nu_i^*=\left(\sum_{i\in\Sigma_0}\mu_i-\frac{\alpha}{2}\right)^+. \tag{43}$$

The extremum solution for any $R\leq R_{\max}$, under equality conditions (35) and (37) is given by

$$D^+(R)=\{\ell_{\max}-\ell_{\min}\}\frac{\alpha}{2}+\sum_{i\in\Sigma}\ell_i\mu_i. \tag{44}$$

Lemma 4.2, Part (b), case 1, characterize the extremum solution for $\sum_{i\in\Sigma_0}\mu_i-(\alpha/2)\geq 0$. Next, the characterization of extremum solution when this condition is violated, that is, when $\sum_{j=1}^k\sum_{i\in\Sigma_{j-1}}\mu_i-(\alpha/2)\leq 0$ for any $k\in\{1,2,\ldots,r\}$, is discussed.

From Lemma 4.2, Part (b), case 2, the lower bound (38), holds with equality if conditions given by (39) are satisfied. Furthermore, from (39b) we have that

$$\sum_{i\in\Sigma_k}\nu_i=\sum_{i\in\Sigma_k}\mu_i-\left(\frac{\alpha}{2}-\sum_{j=1}^k\sum_{i\in\Sigma_{j-1}}\mu_i\right) \tag{45}$$

and conditions $(\alpha/2)-\sum_{j=1}^k\sum_{i\in\Sigma_{j-1}}\mu_i\geq 0$ and (39c) must be satisfied, hence the optimal probability on Σ_k is given by

$$\nu^*(\Sigma_k)\triangleq\sum_{i\in\Sigma_k}\nu_i^*=\left(\sum_{i\in\Sigma_k}\mu_i-\left(\frac{\alpha}{2}-\sum_{j=1}^k\sum_{i\in\Sigma_{j-1}}\mu_i\right)^+\right)^+. \tag{46}$$

The extremum solution for any $R\leq R_{\max}$, under equality conditions (35) and (39) is given by

$$D^+(R)=\sum_{i\in\Sigma}\ell_i\xi_i^+-\sum_{i\in\Sigma}\ell_i\xi_i^-+\sum_{i\in\Sigma}\ell_i\mu_i=\ell_{\max}\left(\frac{\alpha}{2}\right)$$
$$-\ell(\Sigma_k)\left(\frac{\alpha}{2}-\sum_{j=1}^k\sum_{i\in\Sigma_{j-1}}\mu_i\right)+\sum_{j=1}^k\sum_{i\in\Sigma_{j-1}}\ell_i\mu_i+\sum_{i\in\Sigma}\ell_i\mu_i.$$

For $R\in[R_{\max},2]$, Lemma 3.1, part (1), states that $D^+(R)$ is constant. Indeed for $\alpha=\|\xi\|_{TV}=R_{\max}=2(1-\mu(\Sigma^0))$ equality conditions of Lemma 4.2, Part (a), become

$$\sum_{i\in\Sigma^0}\mu_i+\frac{\alpha}{2}=1,\quad \sum_{i\in\Sigma^0}\xi_i^+=\frac{\alpha}{2},\quad \xi_i^+=0 \text{ for } i\in\Sigma\setminus\Sigma^0 \tag{47}$$

and hence

$$\sum_{i\in\Sigma\setminus\Sigma^0}\mu_i-\frac{\alpha}{2}=0,\quad \sum_{i\in\Sigma\Sigma^0}\xi_i^-=\frac{\alpha}{2},\quad \xi_i^-=0 \text{ for } i\in\Sigma^0. \tag{48}$$

Therefore, $\sum_{i\in\Sigma\setminus\Sigma^0}\xi_i^-=\sum_{i\in\Sigma\setminus\Sigma^0}\mu_i$ and hence $\xi_i^-=\mu_i$ for all $i\in\Sigma\setminus\Sigma^0$. The extremum solution for any $R\in[R_{\max},2]$ is given by

$$\begin{aligned}D^+(R)&=\sum_{i\in\Sigma}\ell_i\xi_i^+-\sum_{i\in\Sigma}\ell_i\xi_i^-+\sum_{i\in\Sigma}\ell_i\mu_i\\&\overset{(a)}{=}\sum_{i\in\Sigma^0}\ell_i\xi_i^+-\sum_{i\in\Sigma\setminus\Sigma^0}\ell_i\xi_i^-+\sum_{i\in\Sigma}\ell_i\mu_i\\&=\ell_{\max}\left(\frac{\alpha}{2}\right)-\sum_{i\in\Sigma\setminus\Sigma^0}\ell_i\mu_i+\sum_{i\in\Sigma}\ell_i\mu_i\\&=\ell_{\max}\left(1-\sum_{i\in\Sigma^0}\mu_i\right)+\sum_{i\in\Sigma^0}\ell_i\mu_i=\ell_{\max}.\end{aligned}$$

where (a) follows from (47) and (48). ■

B. Problem 2.2: Finite Alphabet Case

In this subsection we provide the solution of Problem 2.2, by following the procedure utilized to derive the solution of Problem 2.1 (e.g., Section IV-A). The extremum problem is defined by

$$R^-(D)\triangleq\min_{\nu\in\mathbb{Q}(D)}\sum_{i\in\Sigma}|\nu_i-\mu_i| \tag{49}$$

where $\mathbb{Q}(D)\triangleq\{\nu\in\mathcal{M}_1(\Sigma):\sum_{i\in\Sigma}\ell_i\nu_i\leq D\}$.

Define the maximum and minimum values of the sequence by $\ell_{\max}\triangleq\max_{i\in\Sigma}\ell_i$, $\ell_{\min}\triangleq\min_{i\in\Sigma}\ell_i$ and its corresponding support sets by $\Sigma^0\triangleq\{i\in\Sigma:\ell_i=\ell_{\max}\}$, $\Sigma_0\triangleq\{i\in\Sigma:\ell_i=\ell_{\min}\}$. For all remaining sequence, $\{\ell_i:i\in\Sigma\setminus\Sigma^0\cup\Sigma_0\}$, and for $1\leq r\leq|\Sigma\setminus\Sigma^0\cup\Sigma_0|$ define recursively the set of indices for which ℓ achieves its $(k+1)$th largest value by

$$\Sigma^k\triangleq\left\{i\in\Sigma:\ell_i=\max\left\{\ell_\alpha:\alpha\in\Sigma\setminus\Sigma_0\cup\left(\bigcup_{j=1}^k\Sigma^{j-1}\right)\right\}\right\} \tag{50}$$

where $k\in\{1,2,\ldots,r\}$, till all the elements of Σ are exhausted, and define the corresponding maximum value of ℓ on the sequence on these sets by

$$\ell(\Sigma^k)\triangleq\max_{i\in\Sigma\setminus\Sigma_0\cup\left(\bigcup_{j=1}^k\Sigma^{j-1}\right)}\ell_i,\quad k\in\{1,2,\ldots,r\}$$

where r is the number of Σ^k sets which is at most $|\Sigma\setminus\Sigma^0\cup\Sigma_0|$. Clearly, $\ell(\Sigma^1)=\max_{i\in\Sigma\setminus\Sigma^0\cup\Sigma_0}\ell_i$ and so on. Note the analogy between (50) and (31) for Problem 2.1. The main theorem which characterizes the extremum solution of Problem 2.2 is given below.

Theorem 4.3: The solution of the finite alphabet version of Problem 2.2 is given by

$$R^-(D) = \sum_{i\in\Sigma} |\nu_i^* - \mu_i| \tag{51}$$

where the value of $R^-(D)$ is calculated as follows.

1) If

$$D \geq \ell_{\min}\left(\sum_{j=0}^{k}\sum_{i\in\Sigma^j}\mu_i + \sum_{i\in\Sigma_0}\mu_i\right) + \sum_{j=k+1}^{r}\sum_{i\in\Sigma^j}\ell_i\mu_i$$

$$D \leq \ell_{\min}\left(\sum_{j=1}^{k}\sum_{i\in\Sigma^{j-1}}\mu_i + \sum_{i\in\Sigma_0}\mu_i\right) + \sum_{j=k}^{r}\sum_{i\in\Sigma^j}\ell_i\mu_i$$

then

$$R^-(D) = \frac{2\left(D - \ell_{\min}\sum_{i\in\Sigma_0}\mu_i - \ell(\Sigma^k)\sum_{j=1}^{k}\sum_{i\in\Sigma^{j-1}}\mu_i - \sum_{j=k}^{r}\sum_{i\in\Sigma^j}\ell_i\mu_i\right)}{\ell_{\min} - \ell(\Sigma^k)}. \tag{52}$$

2) If $D \geq (\ell_{\min} - \ell_{\max})\sum_{i\in\Sigma^0}\mu_i + \sum_{i\in\Sigma}\ell_i\mu_i$ then

$$R^-(D) = \frac{2\left(D - \sum_{i\in\Sigma}\ell_i\mu_i\right)}{\ell_{\min} - \ell_{\max}}. \tag{53}$$

Moreover, the optimal probabilities are given by

$$\nu^*(\Sigma_0) \triangleq \sum_{i\in\Sigma_0}\nu_i^* = \sum_{i\in\Sigma_0}\mu_i + \frac{\alpha}{2} \tag{54a}$$

$$\nu^*(\Sigma^0) \triangleq \sum_{i\in\Sigma^0}\nu_i^* = \left(\sum_{i\in\Sigma^0}\mu_i - \frac{\alpha}{2}\right)^+ \tag{54b}$$

$$\nu^*(\Sigma^k) \triangleq \sum_{i\in\Sigma^k}\nu_i^* = \left(\sum_{i\in\Sigma^k}\mu_i - \left(\frac{\alpha}{2} - \sum_{j=1}^{k}\sum_{i\in\Sigma^{j-1}}\mu_i\right)^+\right)^+ \tag{54c}$$

$$\alpha = \min\left(R^-(D), 2\left(1 - \sum_{i\in\Sigma_0}\mu_i\right)\right). \tag{54d}$$

where $k = 1, 2, \ldots, r$ and r is the number of Σ^k sets which is at most $|\Sigma \setminus \Sigma^0 \cup \Sigma_0|$.

Proof: For the derivation of Theorem 4.3 see the Appendix. ■

C. Solutions of Related Extremum Problems

In Section II-A we discuss related extremum problems, whose solution can be obtained from those of Problem 2.1 and Problem 2.2. In this Section we give the solution of the finite alphabet version of the related extremum problems described by (8) and (10).

Consider the finite alphabet version of (8), that is

$$R^+(D) \triangleq \sup_{\nu\in\mathcal{M}_1(\Sigma):\sum_{i\in\Sigma}\ell_i\nu_i\leq D} \|\nu - \mu\|_{TV}. \tag{55}$$

As stated in Section II-A, the solution of (55) is obtained from the solution of Problem 2.1, by finding the inverse mapping or by following a similar procedure to the one utilized to derive Theorem 4.3. The main results are stated below.

Theorem 4.4: The solution of the finite alphabet version of (55) is given by

$$R^+(D) = \sum_{i\in\Sigma}|\nu_i^* - \mu_i| \tag{56}$$

where the value of $R^+(D)$ is calculated as follows.

1) If

$$D \geq \ell_{\max}\left(\sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i + \sum_{i\in\Sigma^0}\mu_i\right) + \sum_{j=k}^{r}\sum_{i\in\Sigma_j}\ell_i\mu_i$$

$$D \leq \ell_{\max}\left(\sum_{j=0}^{k}\sum_{i\in\Sigma_j}\mu_i + \sum_{i\in\Sigma^0}\mu_i\right) + \sum_{j=k+1}^{r}\sum_{i\in\Sigma_j}\ell_i\mu_i,$$

then

$$R^+(D) = \frac{2\left(D - \ell_{\max}\sum_{i\in\Sigma^0}\mu_i - \ell(\Sigma_k)\sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i - \sum_{j=k}^{r}\sum_{i\in\Sigma_j}\ell_i\mu_i\right)}{\ell_{\max} - \ell(\Sigma_k)}. \tag{57}$$

2) If $D \leq (\ell_{\max} - \ell_{\min})\sum_{i\in\Sigma_0}\mu_i + \sum_{i\in\Sigma}\ell_i\mu_i$ then

$$R^+(D) = \frac{2\left(D - \sum_{i\in\Sigma}\ell_i\mu_i\right)}{\ell_{\max} - \ell_{\min}}. \tag{58}$$

Moreover, the optimal probabilities are given by

$$\nu^*(\Sigma^0) \triangleq \sum_{i\in\Sigma^0}\nu_i^* = \sum_{i\in\Sigma^0}\mu_i + \frac{\alpha}{2} \tag{59a}$$

$$\nu^*(\Sigma_0) \triangleq \sum_{i\in\Sigma_0}\nu_i^* = \left(\sum_{i\in\Sigma_0}\mu_i - \frac{\alpha}{2}\right)^+ \tag{59b}$$

$$\nu^*(\Sigma_k) \triangleq \sum_{i\in\Sigma_k}\nu_i^* = \left(\sum_{i\in\Sigma_k}\mu_i - \left(\frac{\alpha}{2} - \sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i\right)^+\right)^+ \tag{59c}$$

$$\alpha = \min\left(R^+(D), 2\left(1 - \sum_{i\in\Sigma^0}\mu_i\right)\right) \tag{59d}$$

where, $k = 1, 2, \ldots, r$ and r is the number of Σ_k sets which is at most $|\Sigma \setminus \Sigma^0 \cup \Sigma_0|$.

Consider the finite alphabet version of (10), that is

$$D^-(R) \triangleq \inf_{\nu \in \mathcal{M}_1(\Sigma): \|\nu-\mu\|_{TV} \leq R} \sum_{i \in \Sigma} \ell_i \nu_i. \quad (60)$$

Similarly as before, the solution of (60) is obtained from that of Problem 2.1, but with a reverse computation on the partition of Σ and the mass of the extremum measure on the partition moving in the opposite direction. Below, we give the main theorem.

Theorem 4.5: The solution of the finite alphabet version of (60) is given by

$$D^-(R) = \ell_{\max}\nu^*(\Sigma^0) + \ell_{\min}\nu^*(\Sigma_0) + \sum_{k=1}^{r} \ell(\Sigma^k)\nu^*(\Sigma^k). \quad (61)$$

Moreover, the optimal probabilities are given by

$$\nu^*(\Sigma_0) \triangleq \sum_{i \in \Sigma_0} \nu_i^* = \sum_{i \in \Sigma_0} \mu_i + \frac{\alpha}{2} \quad (62a)$$

$$\nu^*(\Sigma^0) \triangleq \sum_{i \in \Sigma^0} \nu_i^* = \left(\sum_{i \in \Sigma^0} \mu_i - \frac{\alpha}{2}\right)^+ \quad (62b)$$

$$\nu^*(\Sigma^k) \triangleq \sum_{i \in \Sigma^k} \nu_i^* = \left(\sum_{i \in \Sigma^k} \mu_i - \left(\frac{\alpha}{2} - \sum_{j=1}^{k} \sum_{i \in \Sigma^{j-1}} \mu_i\right)^+\right)^+ \quad (62c)$$

$$\alpha = \min(R, R_{\max}), \qquad R_{\max} \triangleq 2\left(1 - \sum_{i \in \Sigma_0} \mu_i\right) \quad (62d)$$

where $k = 1, 2, \ldots, r$ and r is the number of Σ^k sets which is at most $|\Sigma \setminus \Sigma^0 \cup \Sigma_0|$.

Remark 4.6: The statements of Theorems 4.1, 4.3, 4.4, 4.5 are also valid for the countable alphabet case, because their derivations are not restricted to Σ being finite alphabet. For example, $D^+(R)$ is given by

$$D^+(R) = \ell_{\max}\nu^*(\Sigma^0) + \ell_{\min}\nu^*(\Sigma_0) + \sum_{k=1}^{r} \ell(\Sigma_k)\nu^*(\Sigma_k) \quad (63)$$

where the optimal probabilities are given by

$$\nu^*(\Sigma^0) = \mu(\Sigma^0) + \frac{\alpha}{2}, \qquad \nu^*(\Sigma_0) = \left(\mu(\Sigma_0) - \frac{\alpha}{2}\right)^+$$

$$\nu^*(\Sigma_k) = \left(\mu(\Sigma_k) - \left(\frac{\alpha}{2} - \sum_{j=1}^{k} \mu(\Sigma_{j-1})\right)^+\right)^+$$

$$\alpha = \min(R, R_{\max}), \qquad R_{\max} \triangleq 2\left(1 - \mu(\Sigma^0)\right)$$

k is at most countable. It also holds for any $\ell \in BC^+(\Sigma)$ as seen in Section III.

V. Relation of Total Variation Distance to Other Metrics

In this section, we briefly state relations of the total variation distance to other distance metrics, and we give some of its applications.

L_1 Distance Uncertainty

Let $\sigma \in \mathcal{M}_1(\Sigma)$ be a fixed measure (as well as $\mu \in \mathcal{M}_1(\Sigma)$). Define the Radon-Nykodym derivatives $\psi \triangleq d\mu/d\sigma$, $\varphi \triangleq d\nu/d\sigma$ (densities with respect to a fixed $\sigma \in \mathcal{M}_1(\Sigma)$). Then

$$\|\nu - \mu\|_{TV} = \int |\varphi(x) - \psi(x)|\, \sigma(dx).$$

Consider a subset of $\mathbb{B}_R(\mu)$ defined by $\mathbb{B}_{R,\sigma}(\mu) \triangleq \{\nu \in \mathbb{B}_R(\mu) : \nu \ll \sigma, \mu \ll \sigma\} \subseteq \mathbb{B}_R(\mu)$. Then,

$$\mathbb{B}_{R,\sigma}(\mu) = \left\{\varphi \in L_1(\sigma), \varphi \geq 0, \sigma - a.s. : \int_{\Sigma} |\varphi(x) - \psi(x)|\, \sigma(dx) \leq R\right\}.$$

Thus, under the absolute continuity of measures the total variation distance reduces to L_1 distance. Robustness via L_1 distance uncertainty on the space of spectral densities is investigated in the context of Wiener-Kolmogorov theory in an estimation and decision framework in [12], [13]. The extremum problem described under (a) can be applied to abstract formulations of minimax control and estimation, when the nominal system and uncertainty set are described by spectral measures with respect to variation distance.

Relative Entropy Uncertainty Model

Reference [4] The relative entropy of $\nu \in \mathcal{M}_1(\Sigma)$ with respect to $\mu \in \mathcal{M}_1(\Sigma)$ is a mapping $H(\cdot|\cdot) : \mathcal{M}_1(\Sigma) \times \mathcal{M}_1(\Sigma) \longmapsto [0, \infty]$ defined by

$$H(\nu|\mu) \triangleq \begin{cases} \int_{\Sigma} \log\left(\frac{d\nu}{d\mu}\right) d\nu, & \text{if } \nu \ll \mu \\ +\infty, & \text{otherwise.} \end{cases}$$

It is well known that $H(\nu|\mu) \geq 0$, $\forall \nu, \mu \in \mathcal{M}_1(\Sigma)$, while $H(\nu|\mu) = 0 \Leftrightarrow \nu = \mu$. Total variation distance is bounded above by relative entropy via Pinsker's inequality giving

$$\|\nu - \mu\|_{TV} \leq \sqrt{2H(\nu|\mu)},\ \nu, \mu \in \mathcal{M}_1(\Sigma). \quad (64)$$

Given a known or nominal probability measure $\mu \in \mathcal{M}_1(\Sigma)$ the uncertainty set based on relative entropy is defined by $A_{\tilde{R}}(\mu) \triangleq \{\nu \in \mathcal{M}_1(\Sigma) : H(\nu|\mu) \leq \tilde{R}\}$, where $\tilde{R} \in [0, \infty)$. Clearly, the uncertainty set determined by the total variation distance d_{TV}, is larger than that determined by the relative entropy. In other words, for every $r > 0$, in view of Pinsker's inequality (64)

$$\left\{\nu \in \mathcal{M}_1(\Sigma), \nu \ll \mu : H(\nu|\mu) \leq \frac{r^2}{2}\right\} \subseteq \mathbb{B}_R(\mu)$$
$$\equiv \{\nu \in \mathcal{M}_1(\Sigma) : \|\nu - \mu\|_{TV} \leq r\}.$$

Hence, even for those measures which satisfy $\nu \ll \mu$, the uncertainty set described by relative entropy is a subset of the much larger total variation distance uncertainty set. Moreover, by Pinsker's inequality, distance in total variation of probability measures is a lower bound on their relative entropy or Kullback-Leibler distance.

Over the last few years, relative entropy uncertainty model has received particular attention due to various properties (convexity, compact level sets), its simplicity and its connection to risk sensitive pay-off, minimax games, and large deviations [7]–[11]. Recently, an uncertainty model along the spirit of Radon-Nikodym derivative is employed in [26] for portfolio optimization under uncertainty. Unfortunately, relative entropy uncertainty modeling has two disadvantages. 1) It does not define a true metric on the space of measures; 2) relative entropy between two measures is not defined if the measures are not absolutely continuous. The latter rules out the possibility of measures $\nu \in \mathcal{M}_1(\Sigma)$ and $\mu \in \mathcal{M}_1(\Sigma)$, $\tilde{\Sigma} \subset \Sigma$ to be defined on different spaces.[4] It is one of the main disadvantages in employing relative entropy in the context of uncertainty modeling for stochastic controlled diffusions (or SDE's) [27]. Specifically, by invoking a change of measure it can be shown that relative entropy modeling allows uncertainty in the drift coefficient of stochastic controlled diffusions, but not in the diffusion coefficient, because the latter kind of uncertainty leads to measures which are not absolutely continuous with respect to the nominal measure [7].

An anthology of other distances and distance metrics related to total variation distance is found in [4]. In view of the relations between different metrics, such as relative entropy, Kakutani-Hellinger metric, Prohorov's metric, etc, it is clear that the Problem discussed under (1)–(4) gives a sub-optimal solution to the same problem with distance in variation replaced by these metrics.

VI. Simulations

We will illustrate through simple examples how the optimal solution of the different extremum problems behaves. In particular, we present calculations through Section VI-A for maximization problems $D^+(R)$ and $R^+(D)$, when the sequence $\ell = \{\ell_1 \ \ell_2 \ \ldots \ \ell_n\} \in \mathbb{R}^n_+$ consists of a number of ℓ_i's which are equal, and calculations through Section VI-B for the corresponding minimization problems $R^-(D)$ and $D^-(R)$, when the ℓ_i's are not equal.

[4]This corresponds to the case in which the nominal system is a simplified version of the true system and is defined on a lower dimension space.

A. Extremum Problems $D^+(R)$ and $R^+(D)$

Let $\Sigma = \{i : i = 1, 2, \ldots, 8\}$ and for simplicity consider a descending sequence of lengths $\ell = \{\ell \in \mathbb{R}^8_+ : \ell_1 = \ell_2 > \ell_3 = \ell_4 > \ell_5 > \ell_6 = \ell_7 > \ell_8\}$ with corresponding nominal probability vector $\mu \in \mathbb{P}_1(\Sigma)$. Specifically, let $\ell = [1, 1, 0.8, 0.8, 0.6, 0.4, 0.4, 0.2]$, and $\mu = [23/72, 13/72, 10/72, 9/72, 8/72, 4/72, 3/72, 2/72]$. Note that, the sets which correspond to the maximum, minimum and all the remaining lengths are equal to $\Sigma^0 = \{1, 2\}$, $\Sigma_0 = \{8\}$, $\Sigma_1 = \{7, 6\}$, $\Sigma_2 = \{5\}$, $\Sigma_3 = \{4, 3\}$. Fig. 1(a)–(c) depicts the maximum linear functional pay-off subject to total variation constraint, $D^+(R)$, and the optimal probabilities, both given by Theorem 4.1. Fig. 1(b)–(d) depicts the maximum total variation pay-off subject to linear functional constraint, $R^+(D)$, and the optimal probabilities, both given by Theorem 4.4. Recall Lemma 3.1 case 1 and Corollary 3.3. Fig. 1(a) shows that, $D^+(R)$ is a non-decreasing concave function of R and also that is constant in $[R_{\max}, 2]$, where $R_{\max} = 2(1 - \mu(\Sigma^0)) = 1$.

B. Extremum Problems $R^-(D)$ and $D^-(R)$

Let $\Sigma = \{i : i = 1, 2, \ldots, 8\}$ and for simplicity consider a descending sequence of lengths $\ell = \{\ell \in \mathbb{R}^8_+ : \ell_1 > \ell_2 > \ell_3 > \ell_4 > \ell_5 > \ell_6 > \ell_7 > \ell_8\}$ with corresponding nominal probability vector $\mu \in \mathbb{P}_1(\Sigma)$. Specifically, let $\ell = [1, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2]$ and $\mu = [23/72, 13/72, 10/72, 9/72, 8/72, 4/72, 3/72, 2/72]$. Note that, the sets which correspond to the maximum, minimum and all the remaining lengths are equal to $\Sigma^0 = \{1\}$, $\Sigma_0 = \{8\}$, $\Sigma^1 = \{2\}$, $\Sigma^2 = \{3\}$, $\Sigma^3 = \{4\}$, $\Sigma^4 = \{5\}$, $\Sigma^5 = \{6\}$, $\Sigma^6 = \{7\}$. Fig. 2(a)–(c) depicts the minimum total variation pay-off subject to linear functional constraint, $R^-(D)$, and the optimal probabilities, both given by Theorem 4.3. Fig. 2(b)–(d) depicts the minimum linear functional pay-off subject to total variation constraint, $D^-(R)$, and the optimal probabilities, both given by Theorem 4.5. Recall Lemma 3.1 case 2 and Corollary 3.3. Fig. 2(a) shows that, $R^-(D)$ is a non-increasing convex function of D, $D \in [\ell_{\min}, \sum_{i \in \Sigma} \ell_i \mu_i)$. Note that for $D < \ell_{\min} = 0.2$ no solution exists and $R^-(D)$ is zero in $[D_{\max}, \infty)$ where $D_{\max} = \sum_{i=1}^8 \ell_i \mu_i = 0.73$.

VII. Application: Dynamic Programming

Extremum problems to the area of stochastic optimal control are investigated. In particular, we apply the results to minimax dynamic programming, subject to uncertainty of the transition probability distribution.

Consider an inventory control example inspired by [28]. Specifically, an optimal inventory ordering policy of a quantity of a certain item at each of the N periods must be found so as to meet a stochastic demand. Let us denote

- x_k, stock available at the beginning of the kth period;
- u_k, stock ordered at the beginning of the kth period;
- w_k, demand during kth period with given probability distribution;
- h, holding cost per unit item remaining unsold at the end of the kth period;

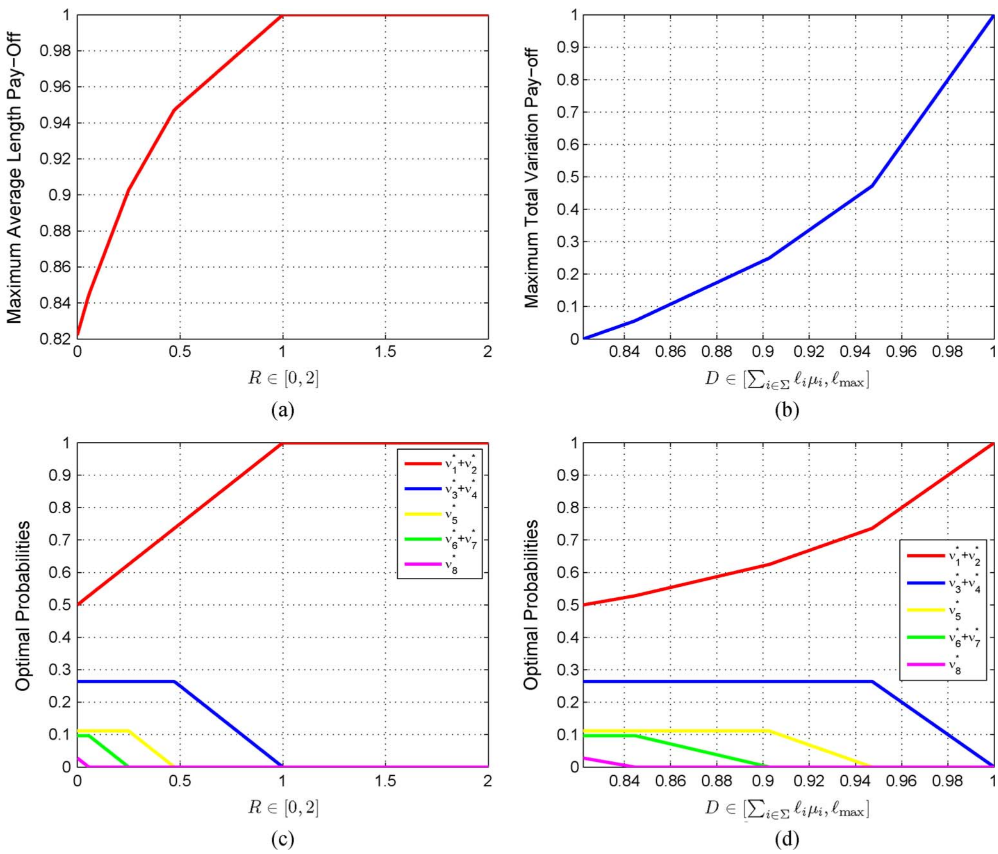

Fig. 1. Example A: optimum solution of (a) $D^+(R)$; and, (b) $R^+(D)$. Optimal probabilities of (c) $D^+(R)$; and, (d) $R^+(D)$.

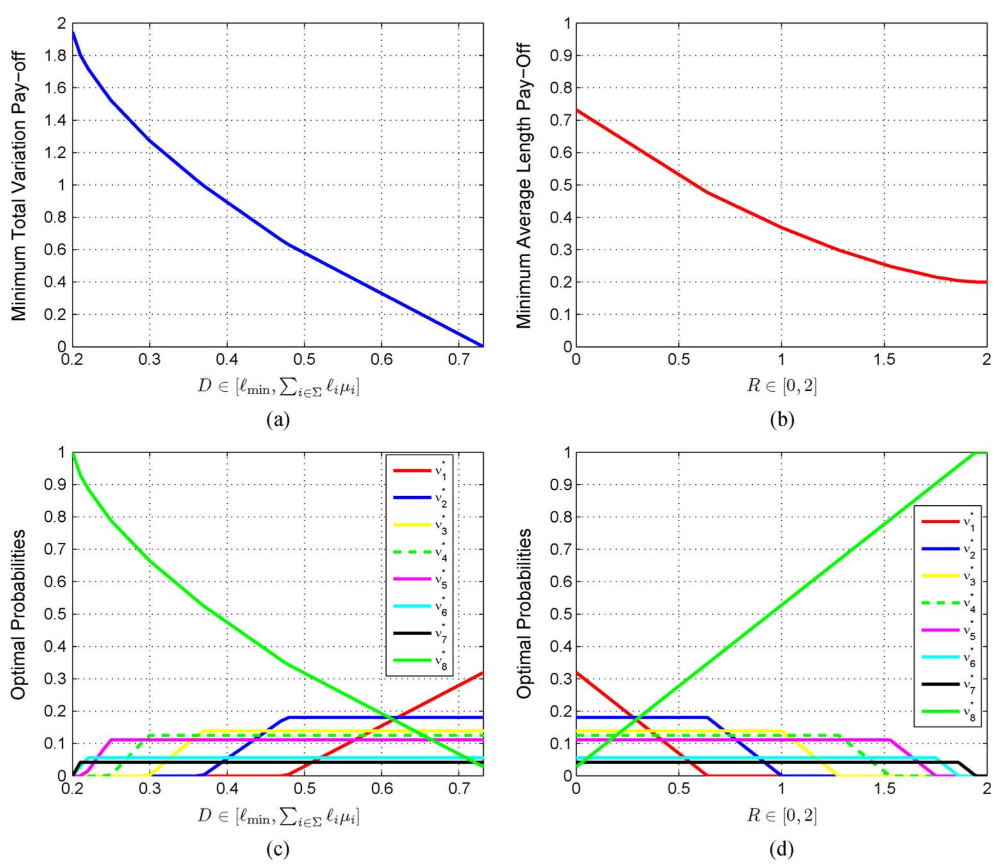

Fig. 2. Example B: optimum solution of (a) $R^-(D)$; and, (b) $D^-(R)$. Optimal probabilities of (c) $R^-(D)$; and, (d) $D^-(R)$.

- c, cost per unit stock ordered;
- p, shortage cost per unit demand unfilled.

The random disturbance at time k, w_k may depend on values of x_k and u_k but not on values of prior disturbances $w_0, \ldots, w_{k-1}$. Excess demand is backlogged and filled as soon as additional inventory becomes available. Inventory and demand are non-negative integers variables. Thus, we assume a nominal system given by

$$x_{k+1} = \max(0, x_k + u_k - w_k) \tag{65}$$

and a total sample pay-off over N periods given by

$$\sum_{k=0}^{N-1} (cu_k + h\max(0, x_k + u_k - w_k) + p\max(0, w_k - x_k - u_k)).$$

We further assume that w_k is independent and identically distributed according to $\mu_{w_k}(\cdot) = \mu_w(\cdot)$. We formulate the problem as a minimax optimization of the expected cost as follows:

$$\min_{u_k \in U_k(x_k)} \max_{\substack{\nu_{w_k}(\cdot): \|\nu_{w_k}(\cdot) - \mu_w(\cdot)\|_{TV} \le R \\ k=0,\ldots,N-1}} \mathbb{E}\left\{ \sum_{k=0}^{N-1} (cu_k + h\max(0, x_k + u_k - w_k) + p\max(0, w_k - x_k - u_k)) \right\}. \tag{66}$$

Assume the following:

- the nominal and the true distribution of $\{w_k : k = 0, 1, \ldots, N-1\}$ is $\mu_{w_k}(\cdot) = \mu_w(\cdot)$, and $\nu_{w_k}(\cdot)$, respectively, $k = 0, 1, \ldots, N-1$;
- the maximum capacity $(x_k + u_k)$ for stock is 2 units;
- the planning horizon $N = 3$ periods;
- the holding cost h and the ordering cost c are both 1 unit;
- the shortage cost p is 3 units;
- the demand w_k has a nominal probability distribution given by, $\mu_w(w_k = 0) = 0.1$, $\mu_w(w_k = 1) = 0.7$, and $\mu_w(w_k = 2) = 0.2$, $k = 0, 1, \ldots, N-1$.

The dynamic programming algorithm for the minimax problem subject to total variation distance uncertainty is given by

$$V_N(x_N) = 0 \tag{67a}$$

$$\begin{aligned} V_k(x_k) &= \min_{0 \le u_k \le 2 - x_k} \max_{\nu_{w_k}(\cdot): \|\nu_{w_k}(\cdot) - \mu_w(\cdot)\|_{TV} \le R} \\ &\quad \mathbb{E}\{u_k + \max(0, x_k + u_k - w_k) + 3\max(0, w_k - x_k - u_k) \\ &\quad + V_{k+1}(\max(0, x_k + u_k - w_k))\} \\ &= \min_{0 \le u_k \le 2 - x_k} \max_{\nu_{w_k}(\cdot): \|\nu_{w_k}(\cdot) - \mu_w(\cdot)\|_{TV} \le R} \mathbb{E}\{\ell_k(x_k, u_k, w_k)\} \\ &\equiv \min_{0 \le u_k \le 2 - x_k} D^+(x_k, u_k, R), \quad k = 0, 1, \ldots, N-1 \end{aligned} \tag{67b}$$

where

$$\ell_k(x_k, u_k, w_k) = u_k + \max(0, x_k + u_k - w_k) + 3\max(0, w_k - x_k - u_k) + V_{k+1}(\max(0, x_k + u_k - w_k)).$$

To address the maximization problem in (67b), for each $k = 0, 1, \ldots, N-1$, $x_k \in \{0, 1, 2\}$ and $0 \le u_k \le 2 - x_k$, define the maximum and minimum values of $\ell(x_k, u_k, w_k)$ by $\ell_{\max}(x_k, u_k) \triangleq \max_{w_k \in \{0,1,2\}} \ell(x_k, u_k, w_k)$ and $\ell_{\min}(x_k, u_k) \triangleq \min_{w_k \in \{0,1,2\}} \ell(x_k, u_k, w_k)$, respectively. Its corresponding support sets are given by

$$\Sigma^0 = \{w_k \in \{0, 1, 2\} : \ell(x_k, u_k, w_k) = \ell_{\max}(x_k, u_k)\}$$

$$\Sigma_0 = \{w_k \in \{0, 1, 2\} : \ell(x_k, u_k, w_k) = \ell_{\min}(x_k, u_k)\}.$$

For all remaining sequence $\{\ell(x_k, u_k, w_k) : w_k \in \{0, 1, 2\} \setminus \Sigma^0 \cup \Sigma_0\}$ and for $1 \le r \le |\{0, 1, 2\} \setminus \Sigma^0 \cup \Sigma_0|$ define recursively the set of indices for which $\ell(x_k, u_k, u_k)$ achieves its $(j+1)$th smallest value by

$$\Sigma_j \triangleq \left\{ w_k \in \{0, 1, 2\} : \ell(x_k, u_k, w_k) = \min\left\{ \ell(x_k, u_k, \alpha_k) : \alpha_k \in \{0, 1, 2\} \setminus \Sigma^0 \cup \left(\bigcup_{i=1}^{j} \Sigma_{i-1} \right) \right\} \right\}, \quad j \in \{1, 2, \ldots, r\}$$

till all the elements of {0, 1, 2} are exhausted. Further, define

$$\ell_{\Sigma_j}(x_k, u_k) \triangleq \min_{w_k \in \{0,1,2\} \setminus \Sigma^0 \cup \left(\bigcup_{i=1}^{j} \Sigma_{i-1}\right)} \ell(x_k, u_k, w_k)$$

where $j \in \{1, 2, \ldots, r\}$. Once we identify the support sets and the corresponding values of the sequence $\ell(x_k, u_k, w_k)$ on these sets, we employ (32), (33) to calculate the maximizing distribution $\nu^*_{w_k}(\cdot)$ and the extremum solution of $D^+(x_k, u_k, R)$. Finally, by employing (67) the optimal cost-to-go and hence the optimal ordering policy are obtained. Alternatively, from the definition of the oscillator seminorm (Remark 3.2, second part), (67) can be expressed as follows:

$$\begin{aligned} V_N(x_N) &= 0 \\ V_k(x_k) &= \min_{0 \le u_k \le 2 - x_k} \Big\{ \mathbb{E}_{\mu_w}\{u_k + \max(0, x_k + u_k - w_k) \\ &\quad + 3\max(0, w_k - x_k - u_k) \\ &\quad + V_{k+1}(\max(0, x_k + u_k - w_k))\} \\ &\quad + \frac{R_k}{2}\Big(\max_{w_k}\{u_k + \max(0, x_k + u_k - w_k) \\ &\quad + 3\max(0, w_k - x_k - u_k) \\ &\quad + V_{k+1}(\max(0, x_k + u_k - w_k))\} \\ &\quad - \min_{w_k}\{u_k + \max(0, x_k + u_k - w_k) \\ &\quad + 3\max(0, w_k - x_k - u_k) \\ &\quad + V_{k+1}(\max(0, x_k + u_k - w_k))\}\Big)\Big\} \end{aligned}$$

TABLE I
DYNAMIC PROGRAMMING ALGORITHM RESULTS

$R=1$			$R=0$		
Stock	Stage.0 Cost-to-go	Stage.0 Optimal Stock to Purchase	Stock	Stage.0 Cost-to-go	Stage.0 Optimal Stock to Purchase
0	7.44	2	0	4.9	1
1	6.44	1	1	3.9	0
2	5.44	0	2	3.35	0
Stock	Stage.1 Cost-to-go	Stage.1 Optimal Stock to Purchase	Stock	Stage.1 Cost-to-go	Stage.1 Optimal Stock to Purchase
0	5.44	2	0	3.3	1
1	4.44	1	1	2.3	0
2	3.44	0	2	1.82	0
Stock	Stage.2 Cost-to-go	Stage.2 Optimal Stock to Purchase	Stock	Stage.2 Cost-to-go	Stage.2 Optimal Stock to Purchase
0	3.2	1	0	1.7	1
1	2.2	0	1	0.7	0
2	1.6	0	2	0.9	0

TABLE II
MAXIMIZING DISTRIBUTION AND SUPPORT SETS FOR $R=1$

	Stock	Optimal Ordering	Support Sets			Maximizing Distribution		
			Σ^0	Σ_0	Σ_1	$\nu^*_{w_k}(\Sigma^0)$	$\nu^*_{w_k}(\Sigma_0)$	$\nu^*_{w_k}(\Sigma_1)$
Stage.0	0	2	{0,1,2}	-	-	1	-	-
	1	1	{0,1,2}	-	-	1	-	-
	2	0	{0,1,2}	-	-	1	-	-
Stage.1	0	2	{0}	{1,2}	-	0.6	0.4	-
	1	1	{0}	{1,2}	-	0.6	0.4	-
	2	0	{0}	{1,2}	-	0.6	0.4	-
Stage.2	0	1	{2}	{1}	{0}	0.7	0.2	0.1
	1	0	{2}	{1}	{0}	0.7	0.2	0.1
	2	0	{0}	{2}	{1}	0.6	0	0.4

where $R_k = R \in [0, 2]$. The problem is solved for two possible values of R for each period resulting in optimal ordering policies as shown in Table I.

By setting $R = 0$, we choose to calculate the optimal control policy, when the true probability distribution $\nu_{w_k}(\cdot) = \mu_w(\cdot)$, $k = 0, 1, 2$. This corresponds to the classical dynamic programming algorithm. From Table I, the resulting optimal ordering policy for each period is to order one unit if the current stock is zero and order nothing otherwise.

By setting the total variation distance $R = 1$, we choose to calculate the optimal control policy, when the true probability distribution is $\nu_{w_k}(\cdot) \neq \mu_w(\cdot)$, $k = 0, 1, 2$. The maximizing distribution $\nu^*_{w_k}(\cdot)$ and its corresponding support sets for each stock available, and the resulting optimal ordering policies at each stage are given in Table II. Taking into consideration the maximization (that is, by setting $R > 0$) the dynamic programming algorithm results in optimal control policies which are more robust with respect to uncertainty, but with the sacrifice of low present and future costs. In cases where the planner needs to balance the desire for low costs with the undesirability of scenarios with high uncertainty, he must choose values of R between 0 and 1. From Table I, the resulting optimal ordering policy for the first two periods is to order two, one and zero units if the current stock is zero, one and two, respectively. For the last period the optimal ordering policy is to order one unit if the current stock is zero and order nothing otherwise.

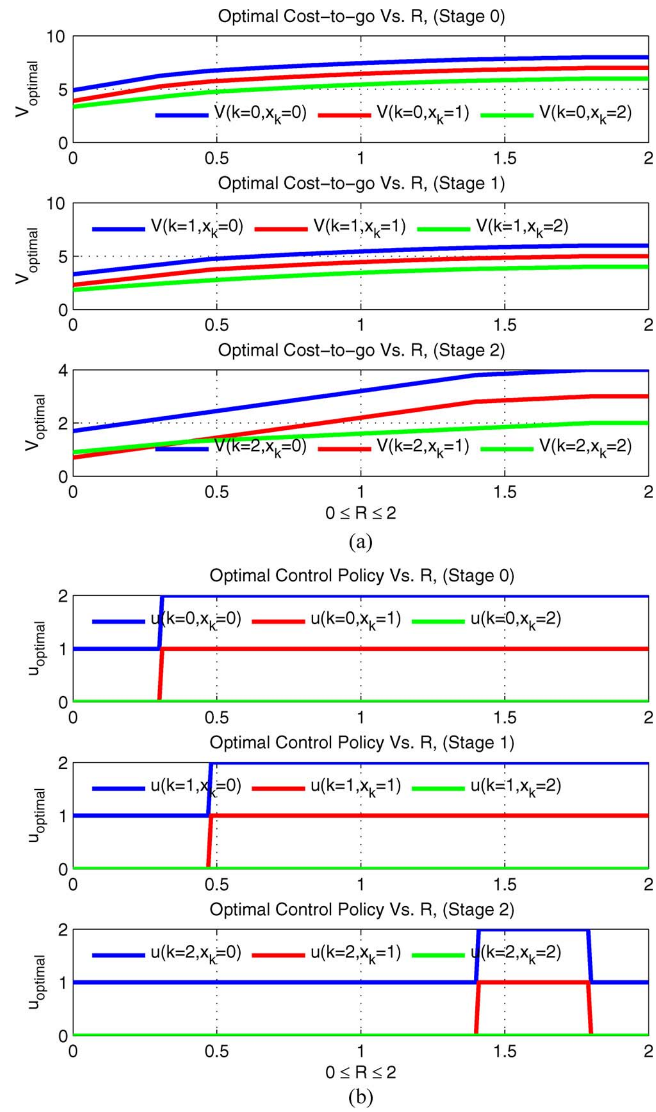

Fig. 3. Section VII: (a) optimal cost-to-go; (b) optimal control policy.

The optimal cost-to-go and the optimal control policy, for each period and for each possible state, as a function of $R \in [0, 2]$, are illustrated in Fig. 3. Clearly, Fig. 3(a) depicts that the optimal cost-to-go is a non-decreasing concave function of R as shown under case 1 in Lemma 3.1.

VIII. CONCLUSION

This paper is concerned with extremum problems involving total variation distance metric as a pay-off subject to linear functional constraints, and vice-versa. These problems are formulated using concepts from signed measures while the theory is developed on abstract spaces. Certain properties and applications of the extremum problems are discussed, while closed form expressions of the extremum measures are derived for finite alphabet spaces. Finally, it is shown through several simulations and an application how the extremum solution of the various problems behaves.

Appendix

Before we proceed with the proof of Theorem 4.3, we give the following Lemma in which lower and upper bounds, which are achievable, are obtained.

Lemma A.1

(a) Lower Bound

$$\sum_{i\in\Sigma} \ell_i \xi_i^+ \geq \ell_{\min}\left(\frac{\alpha}{2}\right). \quad (68)$$

The bound holds with equality if

$$\sum_{i\in\Sigma_0} \mu_i + \frac{\alpha}{2} \leq 1, \qquad \sum_{i\in\Sigma_0} \xi_i^+ = \frac{\alpha}{2}, \qquad \xi_i^+ = 0,\ \forall i \in \Sigma\setminus\Sigma_0. \quad (69)$$

(b) Upper Bound.

Case 1) If $\sum_{i\in\Sigma^0}\mu_i - (\alpha/2) \geq 0$ then

$$\sum_{i\in\Sigma} \ell_i \xi_i^- \leq \ell_{\max}\left(\frac{\alpha}{2}\right). \quad (70)$$

The bound holds with equality if

$$\sum_{i\in\Sigma^0} \mu_i - \frac{\alpha}{2} \geq 0, \qquad \sum_{i\in\Sigma^0} \xi_i^- = \frac{\alpha}{2}, \qquad \xi_i^- = 0,\ \forall i \in \Sigma\setminus\Sigma^0. \quad (71)$$

Case 2) If $\sum_{j=1}^k \sum_{i\in\Sigma^{j-1}}\mu_i - (\alpha/2) \leq 0$ for any $k \in \{1,2,\ldots,r\}$ then

$$\sum_{i\in\Sigma} \ell_i \xi_i^- \leq \ell(\Sigma^k)\left(\frac{\alpha}{2} - \sum_{j=1}^k \sum_{i\in\Sigma^{j-1}} \mu_i\right) + \sum_{j=1}^k \sum_{i\in\Sigma^{j-1}} \ell_i\mu_i. \quad (72)$$

Moreover, equality holds if

$$\sum_{i\in\Sigma^{j-1}} \xi_i^- = \sum_{i\in\Sigma^{j-1}} \mu_i, \text{ for all } j = 1,2,\ldots,k \quad (73a)$$

$$\sum_{i\in\Sigma^k} \xi_i^- = \left(\frac{\alpha}{2} - \sum_{j=1}^k \sum_{i\in\Sigma^{j-1}} \mu_i\right) \quad (73b)$$

$$\sum_{j=0}^k \sum_{i\in\Sigma^j} \mu_i - \frac{\alpha}{2} \geq 0 \quad (73c)$$

$$\xi_i^- = 0 \text{ for all } i \in \Sigma\setminus\Sigma^0\cup\Sigma_1\cup\ldots\cup\Sigma^k. \quad (73d)$$

Proof: Part (a) and Part (b), case 1, follows from Section III-B. The proof of Part (b), case 2, is similar to the proof given for Lemma 4.2, Part (b), case 2, with appropriate changes on Σ^k sets. ■

Proof of Theorem 4.3: From Lemma 3.1, part (2), and Corollary 3.3, we know that for $D \leq D_{\max}$, where $D_{\max} = \sum_{i\in\Sigma}\ell_i\mu_i$, the average constraint holds with equality, that is

$$\sum_{i\in\Sigma}\ell_i\nu_i = \sum_{i\in\Sigma}\ell_i\xi_i^+ - \sum_{i\in\Sigma}\ell_i\xi_i^- + \sum_{i\in\Sigma}\ell_i\mu_i = D.$$

From Lemma A.1, Part (a) and from Part (b), case 1, when equality conditions (69) and (71) are satisfied we have that

$$\ell_{\min}\left(\frac{\alpha}{2}\right) - \ell_{\max}\left(\frac{\alpha}{2}\right) + \sum_{i\in\Sigma}\ell_i\mu_i = D.$$

Solving the above equation with respect to α we get that

$$\alpha = \frac{2\left(D - \sum_{i\in\Sigma}\ell_i\mu_i\right)}{\ell_{\min} - \ell_{\max}}. \quad (74)$$

Since the first equation of (69) is always satisfied, it remains to ensure that the first equation of (71) is also satisfied. By substituting (74) into the first equation of (71) and solving with respect to D we get that if $D \geq (\ell_{\min} - \ell_{\max})\sum_{i\in\Sigma^0}\mu_i + \sum_{i\in\Sigma}\ell_i\mu_i$ then $R^-(D)$ is given by (53). Moreover, the optimal probabilities given by (54a) and (54b) are obtained from the second equation of (69) and (71), respectively.

Lemma A.1, Part (b), case 1, characterize the extremum solution for $\sum_{i\in\Sigma^0}\mu_i - (\alpha/2) \geq 0$. Next, the characterization of extremum solution when this condition is violated, that is, when $\sum_{j=1}^k\sum_{i\in\Sigma^{j-1}}\mu_i - (\alpha/2) \leq 0$ for any $k \in \{1,2,\ldots,r\}$, is discussed.

From Lemma A.1, Part (b), case 2, the upper bound (72), holds with equality if conditions given by (73) are satisfied. Hence,

$$\ell_{\min}\left(\frac{\alpha}{2}\right) - \ell(\Sigma^k)\left(\frac{\alpha}{2} - \sum_{j=1}^k \sum_{i\in\Sigma^{j-1}} \mu_i\right) + \sum_{j=1}^k \sum_{i\in\Sigma^{j-1}} \ell_i\mu_i = D.$$

Solving the above equation with respect to α we get that

$$\alpha = \frac{2\left(D - \ell_{\min}\sum_{i\in\Sigma_0}\mu_i - \ell(\Sigma^k)\sum_{j=1}^k\sum_{i\in\Sigma^{j-1}}\mu_i - \sum_{j=k}^r\sum_{i\in\Sigma^j}\ell_i\mu_i\right)}{\ell_{\min} - \ell(\Sigma^k)}. \quad (75)$$

Substituting (75) into $\sum_{j=1}^k\sum_{i\in\Sigma^{j-1}}\mu_i - (\alpha/2) \leq 0$ and into (73c) and solving with respect to D we get that if

$$D \geq \ell_{\min}\left(\sum_{j=0}^k\sum_{i\in\Sigma^j}\mu_i + \sum_{i\in\Sigma_0}\mu_i\right) + \sum_{j=k+1}^r\sum_{i\in\Sigma^j}\ell_i\mu_i$$

$$D \leq \ell_{\min}\left(\sum_{j=1}^k\sum_{i\in\Sigma^{j-1}}\mu_i + \sum_{i\in\Sigma_0}\mu_i\right) + \sum_{j=k}^r\sum_{i\in\Sigma^j}\ell_i\mu_i,$$

then $R^-(D)$ is given by (52). Moreover, the optimal probability on Σ^k given by (54c) is obtained from (73b).

For $D \in [D_{\max}, \infty)$, it is straightforward that the extremum measure is given by $\nu^* = \mu$ and hence $R^-(D) = 0$. ■

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 1991.

[2] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. London, U.K.: Springer-Verlag, 1993.

[3] A. L. Gibbs and F. E. SU, "On choosing and bounding probability metrics," *Int. Statist. Rev.*, vol. 70, no. 3, pp. 419–435, Dec. 2002.

[4] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. New York, NY, USA: Wiley, 1997.

[5] C. D. Charalambous, I. Tzortzis, and F. Rezaei, "Stochastic optimal control of discrete-time systems subject to conditional distribution uncertainty," in *Proc. 50th IEEE Conf. Decision and Control and European Control Conference*, Orlando, FL, USA, Dec. 12–15, 2011, pp. 6407–6412.

[6] N. Dunford and J. Schwartz, *Linear Operators: Part 1: General Theory*. New York, NY, USA: Interscience, 1957.

[7] P. D. Pra, L. Meneghini, and W. J. Runggaldier, "Connections between stochastic control and dynamic games," *Math. Control Signals Syst.*, vol. 9, no. 4, pp. 303–326, 1996.

[8] V. A. Ugrinovskii and I. R. Petersen, "Finite horizon minimax optimal control of stochastic partially observed time varying uncertain systems," *Math. Control Signals Syst.*, vol. 12, no. 1, pp. 1–23, 1999.

[9] I. R. Petersen, M. R. James, and P. Dupuis, "Minimax optimal control of stochastic uncertain systems with relative entropy constraints," *IEEE Trans. Autom. Control*, vol. 45, no. 3, pp. 398–412, Mar. 2000.

[10] N. U. Ahmed and C. D. Charalambous, "Minimax games for stochastic systems subject to relative entropy uncertainty: Applications to sde's on Hilbert spaces," *J. Math. Control Signals Syst.*, vol. 19, no. 1, pp. 65–91, Feb. 2007.

[11] C. D. Charalambous and F. Rezaei, "Stochastic uncertain systems subject to relative entropy constraints: Induced norms and monotonicity properties of minimax games," *IEEE Trans. Autom. Control*, vol. 52, no. 4, pp. 647–663, Apr. 2007.

[12] H. V. Poor, "On robust Wiener filtering," *IEEE Trans. Autom. Control*, vol. 25, no. 3, pp. 531–536, Jun. 1980.

[13] K. S. Vastola and H. V. Poor, "On robust Wiener-Kolmogorov theory," *IEEE Trans. Inform. Theory*, vol. 30, no. 2, pp. 315–327, Mar. 1984.

[14] E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems*. New York, NY, USA: Springer-Verlag, 1985.

[15] A. Ferrante, M. Pavon, and F. Ramponi, "Hellinger vs. Kullback-Leibler multivariable spectrum approximation," *IEEE Trans. Autom. Control*, vol. 53, no. 5, pp. 954–967, May 2008.

[16] T. T. Georgiou, "Relative entropy and the multivariable multidimensional moment problem," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1052–1066, Mar. 2006.

[17] A. Ferrante, M. Pavon, and F. Ramponi, "Constrained approximation in the Hellinger distance," in *Proc. Eur. Control Conf.*, Kos, Greece, Jul. 2–5, 2007, pp. 322–327.

[18] T. T. Georgiou and A. Lindquist, "Kullback-Leibler approximation of spectral density functions," *IEEE Trans. Inform. Theory*, vol. 49, no. 11, pp. 2910–2917, Nov. 2003.

[19] M. Pavon and A. Ferrante, "On the Georgiou-Lindquist approach to constrained Kullback-Leibler approximation of spectral densities," *IEEE Trans. Autom. Control*, vol. 51, no. 4, pp. 639–644, Apr. 2006.

[20] F. Rezaei, C. D. Charalambous, and N. U. Ahmed, "Optimal control of uncertain stochastic systems subject to total variation distance uncertainty," *SIAM J. Control and Optimiz.*, vol. 50, no. 5, pp. 2683–2725, Sep. 2012.

[21] E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev.*, vol. 106, pp. 620–630, 1957.

[22] E. T. Jaynes, "Information theory and statistical mechanics ii," *Phys. Rev.*, vol. 108, pp. 171–190, 1957.

[23] J. S. Baras and M. Rabi, "Maximum entropy models, dynamic games, and robust output feedback control for automata," in *Proc. 44th IEEE Conf. Decision and Control, and the European Control Conf.*, Seville, Spain, Dec. 12–15, 2005.

[24] F. Rezaei, C. D. Charalambous, and N. U. Ahmed, *Optimization of Stochastic Uncertain Systems: Entropy Rate Functionals, Minimax Games and Robustness. A Festschrift in Honor of Robert J Elliott*. Singapore: World Scientific, 2012, ser. Advances in Statistics, Probability and Actuarial Science.

[25] P. R. Halmos, *Measure Theory*. New York, NY, USA: Springer-Verlag, 1974.

[26] B. Oksendal and A. Sulem, "Portfolio Optimization Under Model Uncertainty and bsde Games," Institut National de Recherche en Informatique et en Automatique, Rapport de Recherche 7554, 2011.

[27] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. San Francisco, CA, USA: Holden-Day, 1964.

[28] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Belmont, MA, USA: Athena Scientific, 2005.

Charalambos D. Charalambous received the B.S., M.E., and Ph.D. degrees in 1987, 1988, and 1992, respectively, all from the Department of Electrical Engineering, Old Dominion University, VA, USA.

In 2003, he joined the Department of Electrical and Computer Engineering, University of Cyprus, where he served as Associate Dean of the School of Engineering until 2009. He was an Associate Professor at the University of Ottawa, School of Information Technology and Engineering, from 1999 to 2003. He served as a non-tenure faculty member on the faculty of the Department of Electrical and Computer Engineering, McGill University, from 1995 to 1999. From 1993 to 1995, he was a Postdoctoral Fellow at Idaho State University. His research group ICCC SystemS, Information, Communication and Control of Complex Systems, is interested in stochastic dynamical decision and control systems, information theory and its applications in telecommunication systems, optimization subject to ambiguity, stochastic dynamic games, large scale distributed and decentralized decision systems, and mathematical finance.

Dr. Charalambous is currently an Associate Editor of the *Systems and Control Letters* and *Mathematics of Control, Signals, and Systems*. In the past he served as an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL and IEEE COMMUNICATIONS LETTERS. In 2001, he received the Premier's Research Excellence Award of the Ontario Province of Canada.

Ioannis Tzortzis received the Diploma degree from the Higher Technical Institute, Nicosia, Cyprus, in 2003, the B.Sc. degree from Budapest University of Technology and Economics, Budapest, Hungary, in 2007, and the M.Sc. degree from the University of Cyprus, Nicosia, Cyprus, in 2009, all in electrical and electronic engineering. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Cyprus.

His research interests include stochastic processes and systems, minimax dynamic games, model reduction, and control and optimization.

Sergey Loyka was born in Minsk, Belarus. He received the Ph.D. degree in radio engineering from the Belorussian State University of Informatics and Radioelectronics (BSUIR), Minsk, Belarus, in 1995 and the M.S. degree (Hons.) from Minsk Radioengineering Institute, Minsk, Belarus, in 1992.

Since 2001, he has been a faculty member at the School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON, Canada. Prior to that, he was a Research Fellow in the Laboratory of Communications and Integrated Microelectronics (LACIME) of Ecole de Technologie Superieure, Montreal, QC, Canada; a Senior Scientist at the Electromagnetic Compatibility Laboratory of BSUIR, Belarus; and an Invited Scientist at the Laboratory of Electromagnetism and Acoustic (LEMA), Swiss Federal Institute of Technology, Lausanne, Switzerland. His areas of research are wireless communications and networks and, in particular, MIMO systems and security aspects of such systems, in which he has published extensively.

Dr. Loyka received a number of awards from the URSI, the IEEE, the Swiss, Belarus and former USSR governments, and the Soros Foundation.

Themistoklis Charalambous received the B.A. and M.Eng. degrees in electrical and information sciences from Trinity College, Cambridge University, Cambridge, U.K., and the Ph.D. degree from the Control Laboratory, Engineering Department, Cambridge University.

Following this, he joined the Human Robotics Group as a Research Associate at Imperial College London and worked as a Visiting Lecturer at the Department of Electrical and Computer Engineering, University of Cyprus. He is currently a Research Associate with the Department of Automatic Control of the School of Electrical Engineering, Royal Institute of Technology (KTH). His research involves cooperative control, distributed decision making, game theory, and control to various resource allocation problems in complex and networked systems.