Extremum Problems with Total Variation Distance

Charalambos D. Charalambous[1], Ioannis Tzortzis[2], Sergey Loyka[3] and Themistoklis Charalambous[4]

Abstract— **The aim of this paper is to investigate extremum problems with pay-off the total variational distance metric subject to linear functional constraints both defined on the space of probability measures, as well as related problems. 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, while the high level ideas are also discussed for denumerable spaces endowed with the discrete topology.**

I. INTRODUCTION

Total variational 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]. Moreover, it is used to define the Dobrushin coefficient when studying the contraction property of transition probability distributions, and in showing ergodicity of Markov chains [2]. Distance in total variation of probability measure is related via upper and lower bounds to an anthology of distances and distance metrics.

In minimax formulations of robust uncertain stochastic systems described by probability measures, it is often desirable to identify proper models of probabilistic uncertainty to quantify the proximity of the probability measures induced by the true and the nominal systems, via distance metrics.

In this paper, we formulate and solve the following extremum problems involving the total variational distance metric: (a) extremum problems of linear functionals defined on the space of measures subject to total variational distance metric constraints; (b) extremum problems of total variational distance metric subject to linear functional constraints defined on the space of measures. We also give examples to illustrate the behavior of the extremum solutions. In particular, three such examples which validate our results can be found in Section VI. Through the first two examples, the maximization problems described under (a) are examined. The optimal solution of such problems as a function of total variational distance is found to be non-decreasing and concave. Through the last example, the minimization problems described under (b) are examined. The optimal solution of such problems as a function of linear constraints is found to be non-increasing and convex. Similar results can be obtained for the minimization problems described by (a), as well as for maximization problems described by (b). The formulation of these extremum problems, their applications, and contributions of this paper are developed at abstract level, in which systems are represented by probability distributions on abstract spaces (complete separable metric space, known as Polish spaces [3]), 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.

The exposition of the paper is organized as follows. In Section II, total variational distance is defined, followed by its relations to other distance metrics. In Section III, the extremum problems are introduced, while several related problems are discussed. In Section IV and V, signed measures are utilized to characterize the extremum measure on abstract and on finite alphabet spaces, respectively. Finally, in Section VI several examples are worked out to illustrate the theoretical results obtained in preceding sections.

II. MOTIVATION FOR TOTAL VARIATIONAL DISTANCE

In this section, we introduce the total variational distance and some of its relations to other distance metrics.
Let (Σ, d_Σ) denote a complete, separable metric space and $(\Sigma, \mathscr{B}(\Sigma))$ the corresponding measurable space, when $\mathscr{B}(\Sigma)$ is the σ-algebra generated by open sets in Σ. Let $\mathscr{M}_1(\Sigma)$ denote the set of probability measures on $\mathscr{B}(\Sigma)$.

Total Variational Distance. The total variational distance[1] is a metric $d_{TV} : \mathscr{M}_1(\Sigma) \times \mathscr{M}_1(\Sigma) \to [0,\infty)$ defined by

$$d_{TV}(\alpha,\beta) \equiv ||\alpha-\beta||_{TV} \stackrel{\triangle}{=} \sup_{P\in\mathscr{P}(\Sigma)} \sum_{F_i\in P} |\alpha(F_i)-\beta(F_i)|,$$

where $\alpha,\beta \in \mathscr{M}_1(\Sigma)$ and $\mathscr{P}(\Sigma)$ denotes the collection of all finite partitions of Σ. With respect to this metric $(\mathscr{M}_1(\Sigma), d_{TV})$ is a complete metric space. Moreover, since the elements of $\mathscr{M}_1(\Sigma)$ 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.

[1]C. D. Charalambous is with the Faculty of Electrical Engineering, University of Cyprus, Nicosia, Cyprus chadcha@ucy.ac.cy
[2]I. Tzortzis is with the Electrical Engineering, University of Cyprus, Nicosia, Cyprus tzortzis.ioannis@ucy.ac.cy
[3]S. Loyka is with the School of Electrical Engineering and Computer Science, University of Ottawa, Canada sloyka@site.uottawa.ca
[4]T. Charalambous is with the School of Electrical Engineering, Royal Institute of Technology, Stockholm, Sweden themisc@kth.se

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

Given a known or nominal probability measure $\mu \in \mathscr{M}_1(\Sigma)$, one can quantify the uncertainty set via the ball with respect to the variational distance, centered at the nominal measure $\mu \in \mathscr{M}_1(\Sigma)$, having radius $R \in [0,2]$, by the set

$$\mathbb{B}_R(\mu) \stackrel{\triangle}{=} \Big\{ \nu \in \mathscr{M}_1(\Sigma) : ||\nu - \mu||_{TV} \leq R \Big\}. \tag{1}$$

Quantifying uncertainty via the metric $||\cdot||_{TV}$ does not require absolute continuity of measures[2], e.g., 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), variational distance uncertainty set models the case when both the drift and diffusion coefficients of SDE's are unknown.

Distance in variation reduces, under certain conditions, to L_1 distance. Specifically, if ν, μ are absolutely continuous with respect to a fixed measure σ, say $\psi \stackrel{\triangle}{=} \frac{d\mu}{d\sigma}$, $\phi \stackrel{\triangle}{=} \frac{d\nu}{d\sigma}$ then

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

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 [4], [5]. 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 variational distance.

Relative Entropy Uncertainty Model. The relative entropy of $\nu \in \mathscr{M}_1(\Sigma)$ with respect to $\mu \in \mathscr{M}_1(\Sigma)$ is a mapping $H(\cdot|\cdot) : \mathscr{M}_1(\Sigma) \times \mathscr{M}_1(\Sigma) \mapsto [0,\infty]$ defined by

$$H(\nu|\mu) \stackrel{\triangle}{=} \begin{cases} \int_\Sigma \log(\frac{d\nu}{d\mu}) d\nu & \text{if } \nu << \mu \\ +\infty & \text{otherwise} \end{cases}$$

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

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

Given a known or nominal probability measure $\mu \in \mathscr{M}_1(\Sigma)$ the uncertainty set based on relative entropy is defined by $A_{\tilde{R}}(\mu) \stackrel{\triangle}{=} \big\{ \nu \in \mathscr{M}_1(\Sigma) : H(\nu|\mu) \leq \tilde{R} \big\}$, 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 (2):

$$\Big\{ \nu \in \mathscr{M}_1(\Sigma), \nu << \mu : H(\nu|\mu) \leq \frac{r^2}{2} \Big\} \subseteq \mathbb{B}_r(\mu).$$

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 $\nu \in \mathscr{M}_1(\Sigma)$ and $\mu \in \mathscr{M}_1(\Sigma)$, $\tilde{\Sigma} \subset \Sigma$ to be defined on different spaces[3].

III. Extremum Problems

Suppose the probability measure $\nu \in \mathscr{M}_1(\Sigma)$ is unknown, while modeling techniques give access to a nominal probability measure $\mu \in \mathscr{M}_1(\Sigma)$. Having constructed the nominal probability measure, one may construct from empirical data, the distance of the two measures with respect to the total variational 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 $\nu \in \mathscr{M}_1(\Sigma)$, centered at the nominal distribution $\mu \in \mathscr{M}_1(\Sigma)$, and lying within the ball of radius R, with respect to the total variational distance $||\cdot||_{TV}$. The precise problem investigated is stated below. Define the spaces

$$BC(\Sigma) \stackrel{\triangle}{=} \{ \ell : \Sigma \mapsto \mathbb{R} : \ell \text{ are bounded continuous} \},$$

$$BM(\Sigma) \stackrel{\triangle}{=} \{ \ell : \Sigma \mapsto \mathbb{R} : \ell \text{ are bounded measurable functions} \}.$$

When endowed with the sup norm $||\ell|| \stackrel{\triangle}{=} \sup_{x \in \Sigma} |\ell(x)|$, then $BC(\Sigma)$ and $BM(\Sigma)$ are Banach spaces. Let $BC^+(\Sigma)$, $BM^+(\Sigma)$ denote subsets of $BC(\Sigma)$, $BM(\Sigma)$, with ℓ non-negative, respectively.

Problem III.1 *Given a fixed nominal distribution $\mu \in \mathscr{M}_1(\Sigma)$ and a parameter $R \in [0,2]$, define the class of true distributions by*

$$\mathbb{B}_R(\mu) \stackrel{\triangle}{=} \Big\{ \nu \in \mathscr{M}_1(\Sigma) : ||\nu - \mu||_{TV} \leq R \Big\}, \tag{3}$$

and the average pay-off with respect to the true probability measure $\nu \in \mathbb{B}_R(\mu) \subset \mathscr{M}_1(\Sigma)$ by

$$\mathbb{L}_1(\nu) \stackrel{\triangle}{=} \int_\Sigma \ell(x) \nu(dx), \qquad \ell \in BC^+(\Sigma) \ \ or \ \ BM^+(\Sigma). \tag{4}$$

The objective is to solve the extremum problem

$$D^+(R) \stackrel{\triangle}{=} \sup_{\nu \in \mathbb{B}_R(\mu)} \int_\Sigma \ell(x) \nu(dx), \quad \forall R \in [0,2]. \tag{5}$$

Problem III.1 is a convex optimization problem on the space of probability measures. It is discussed in [6] in the context of stochastic optimal control. Note that, $BC(\Sigma)$ can be generalized to $L^{\infty,+}(\Sigma, \mathscr{B}(\Sigma), \nu)$, the set of all $\mathscr{B}(\Sigma)$-measurable, non-negative essentially bounded functions defined $\nu - a.e.$ endowed with the essential supremum norm. In the context of minimax theory, Problem III.1 is important in uncertain stochastic control, estimation, and decision, formulated via minimax optimization. Such formulations are found in [7] utilizing relative entropy uncertainty, and in [4] utilizing L_1 distance uncertainty. The second extremum problem is defined below.

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

[3] 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.

Problem III.2 *Given a fixed nominal distribution $\mu \in \mathscr{M}_1(\Sigma)$ and a parameter $D \in [0,\infty)$, define the class of true distributions by*

$$\mathbb{Q}(D) \stackrel{\triangle}{=} \Big\{ \nu \in \mathscr{M}_1(\Sigma) : \int_\Sigma \ell(x)\nu(dx) \le D \Big\}, \tag{6}$$

where $\ell \in BC^+(\Sigma)$ or $\ell \in BM^+(\Sigma)$. The total variation pay-off with respect to the true probability measure $\nu \in \mathbb{Q}(D) \subset \mathscr{M}_1(\Sigma)$ is defined by

$$\mathbb{L}_2(\nu) \stackrel{\triangle}{=} ||\nu - \mu||_{TV}. \tag{7}$$

The objective is to solve the extremum problem

$$R^-(D) \stackrel{\triangle}{=} \inf_{\nu \in \mathbb{Q}(D)} ||\nu - \mu||_{TV}, \tag{8}$$

whenever [4] $\int_\Sigma \ell(x)\mu(dx) > D$.

Problem III.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. Problem III.2 is also important in spectral measure or density approximation as follows. Recall that a function $\{R(\tau) : -\infty \le \tau \le \infty\}$ is the covariance function of a quadratic mean continuous and wide-sense stationary process if and only if it is of the form [8]

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

where $F(\cdot)$ is a finite Borel measure on $\mathbb{R}$, called spectral measure. Thus, by proper normalization of $F(\cdot)$ via $F_N(d\nu) \stackrel{\triangle}{=} \frac{1}{R(0)}F(d\nu)$, then $F_N(d\nu)$ is a probability measure on $\mathscr{B}(\mathbb{R})$, and hence Problem III.2 can be used to approximate the class of spectral measures which satisfy moment estimates.

A. Related Extremum Problems

Problems III.1, III.2 are related to additional extremum problems which are introduced below.

(1) The solution of Problem III.1 gives the solution of the problem defined by

$$R^+(D) \stackrel{\triangle}{=} \sup_{\nu \in \mathscr{M}_1(\Sigma): \int_\Sigma \ell(x)\nu(dx) \le D} ||\nu - \mu||_{TV}. \tag{9}$$

Specifically, $R^+(D)$ is the inverse mapping of $D^+(R)$.

(2) Let ν and μ be absolutely continuous with respect to the Lebesgue measure so that $\varphi(x) \stackrel{\triangle}{=} \frac{d\nu}{dx}(x)$, $\psi(x) \stackrel{\triangle}{=} \frac{d\mu}{dx}(x)$ (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, Problem III.1 and Problem (9) 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 $\mathscr{M}_1(\Sigma)$ identified with the standard probability simplex in $\mathbb{R}^{|\Sigma|}$, that is, the set of all $|\Sigma|$-dimensional vectors which are probability vectors, and $\ell(x) \stackrel{\triangle}{=} -\log \nu(x), x \in \Sigma$, where $\{\nu(x) : x \in \Sigma\} \in \mathscr{M}_1(\Sigma)$, $\{\mu(x) : x \in \Sigma\} \in \mathscr{M}_1(\Sigma)$. Then Problem III.1 is equivalent to maximizing the entropy of $\{\nu(x) : x \in \Sigma\}$ subject to total variational distance metric constraint defined by

$$D^+(R) = \max_{\nu \in \mathscr{M}_1(\Sigma): \sum_{x\in\Sigma} |\nu(x) - \mu(x)| \le R} H(\nu). \tag{10}$$

Problem (10) is of interest when the concept of insufficient reasoning (e.g., Jayne's maximum entropy principle) is applied to construct a model for $\nu \in \mathscr{M}_1(\Sigma)$, subject to information quantified via total variational distance metric between ν and an empirical distribution μ.

(4) Similarly as in (1), Problem III.2 gives the solution of the problem defined by

$$D^-(R) \stackrel{\triangle}{=} \inf_{\nu \in \mathscr{M}_1(\Sigma): ||\nu-\mu||_{TV} \le R} \int_\Sigma \ell(x)\nu(dx). \tag{11}$$

Problem III.2 and Problem (11) are important in approximating a class of probability distributions or spectral measures by reduced ones.

IV. Characterization of Extremum Measure on Abstract Spaces

This section utilizes signed measures and some of their properties to find the solution of Problem III.1. First, we describe some properties of Problems III.1 and III.2. Similar properties hold for the other problems.

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

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

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

$$R^-(D) = \inf_{\int_\Sigma \ell(x)\nu(dx) = D} \mathbb{L}_2(\nu), \quad \text{if} \quad D \le D_{\max}, \tag{13}$$

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. See [9]. ■

Let $\mathscr{M}_{sm}(\Sigma)$ denote the set of finite signed measures. Then, any $\eta \in \mathscr{M}_{sm}(\Sigma)$ has a Jordan decomposition [5] $\{\eta^+, \eta^-\}$ such that $\eta = \eta^+ - \eta^-$, and the total variation of η is defined by $||\eta||_{TV} \stackrel{\triangle}{=} \eta^+(\Sigma) + \eta^-(\Sigma)$. Define the following subset $\mathbb{M}_0(\Sigma) \stackrel{\triangle}{=} \Big\{\eta \in \mathscr{M}_{sm}(\Sigma) : \eta(\Sigma) = 0\Big\}$. If $\xi \in \mathbb{M}_0(\Sigma)$, then $\xi(\Sigma) = 0$, which implies that $\xi^+(\Sigma) = \xi^-(\Sigma)$, and hence $\xi^+(\Sigma) = \xi^-(\Sigma) = \frac{||\xi||_{TV}}{2}$. Then, $\xi \stackrel{\triangle}{=} \nu - \mu \in \mathbb{M}_0(\Sigma)$ and hence $\xi = (\nu-\mu)^+ - (\nu-\mu)^- \equiv \xi^+ - \xi^-$.

[4] If $\int_\Sigma \ell(x)\mu(dx) \le D$ then $\nu^* = \mu$ is the trivial extremum measure of (8).

Consider the pay-off of Problem III.1, for $\ell \in BC^+(\Sigma)$. Then the following inequalities hold.

$$\begin{aligned}
\mathbb{L}_1(\nu) &\overset{(a)}{=} \int_\Sigma \ell(x)\left(\xi^+(dx)-\xi^-(dx)\right)+\int_\Sigma \ell(x)\mu(dx)\\
&\overset{(b)}{\leq} \sup_{x\in\Sigma}\ell(x)\xi^+(\Sigma)-\inf_{x\in\Sigma}\ell(x)\xi^-(\Sigma)+\int_\Sigma \ell(x)\mu(dx)\\
&\overset{(c)}{=} \sup_{x\in\Sigma}\ell(x)\frac{||\xi||_{TV}}{2}-\inf_{x\in\Sigma}\ell(x)\frac{||\xi||_{TV}}{2}+\int_\Sigma \ell(x)\mu(dx)\\
&= \left\{\sup_{x\in\Sigma}\ell(x)-\inf_{x\in\Sigma}\ell(x)\right\}\frac{||\xi||_{TV}}{2}+\int_\Sigma \ell(x)\mu(dx), \quad (14)
\end{aligned}$$

where (a) follows by adding and subtracting $\int \ell d\mu$, and from Jordan decomposition, (b) follows due to $\ell\in BC^+(\Sigma)$, (c) follows because any $\xi\in\mathbb{M}_0(\Sigma)$ satisfies $\xi^+(\Sigma)=\xi^-(\Sigma)=\frac{1}{2}||\xi||_{TV}$. For a given $\mu\in\mathscr{M}_1(\Sigma)$ and $\nu\in\mathbb{B}_R(\mu)$ define the set

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

The upper bound in the right hand side of (14) is achieved by $\xi^*\in\widetilde{\mathbb{B}}_R(\mu)$ as follows. Let

$$\begin{aligned}
x^0\in\Sigma^0 &\overset{\triangle}{=} \left\{x\in\Sigma:\ell(x)=\sup\{\ell(x):x\in\Sigma\}\overset{\triangle}{=}M\right\},\\
x_0\in\Sigma_0 &\overset{\triangle}{=} \left\{x\in\Sigma:\ell(x)=\inf\{\ell(x):x\in\Sigma\}\overset{\triangle}{=}m\right\}.
\end{aligned}$$

Take

$$\xi^*(dx)=\nu^*(dx)-\mu(dx)=\frac{R}{2}\left(\delta_{x^0}(dx)-\delta_{x_0}(dx)\right), \quad (15)$$

where $\delta_y(dx)$ denotes the Dirac measure concentrated at $y\in\Sigma$. This is indeed a signed measure with $||\xi||_{TV}=||\nu^*-\mu||_{TV}=R$, and $\int_\Sigma \ell(x)(\nu^*-\mu)(dx)=\frac{R}{2}(M-m)$. Hence, by using (15) as a candidate of the maximizing distribution, then Problem III.1 is equivalent to

$$\int_\Sigma \ell(x)\nu^*(dx)=\frac{R}{2}\left\{\sup_{x\in\Sigma}\ell(x)-\inf_{x\in\Sigma}\ell(x)\right\}+\int_\Sigma \ell(x)\mu(dx), \quad (16)$$

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 $A\in\mathscr{B}(\Sigma)$. Alternatively, the pay-off $\mathbb{L}_1(\nu^*)$ can be written as

$$\mathbb{L}_1(\nu^*)=\int_{\Sigma^0}M\nu^*(dx)+\int_{\Sigma_0}m\nu^*(dx)+\int_{\Sigma\setminus\Sigma^0\cup\Sigma_0}\ell(x)\mu(dx). \quad (17)$$

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

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

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

$$\nu^*(A)=\mu(A), \quad \forall A\subseteq\Sigma\setminus\Sigma^0\cup\Sigma_0. \quad (18c)$$

For any $R\in[0,2]$ such that $\nu^*(\Sigma^0)<1$ and $\nu^*(\Sigma_0)>0$, then (18) is the maximizing distribution while the resulting pay-off is (17). When these conditions are violated the measure ν^* on the sets Σ^0, Σ_0 and $\Sigma\setminus\Sigma^0\cup\Sigma_0$ remains to be identified so the maximizing measure ν^* is characterized for all $R\in[0,2]$. The complete characterization of the extremum measure ν^* will be given in the next section building on the discussion of this section. Similar results are obtained for Problem III.2.

V. Characterization of Extremum Measures For Finite Alphabets

In this section, we discuss the case when Σ is a finite alphabet space to give the intuition into the solution procedure. Consider the finite alphabet case $(\Sigma,\mathscr{M})$, where $\text{card}(\Sigma)=|\Sigma|$ is finite, $\mathscr{M}=2^{|\Sigma|}$. Thus, ν and μ are point mass distributions on Σ. Define the set of probability vectors on Σ by

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

Thus, $p\in\mathbb{P}(\Sigma)$ is a probability vector in $\mathbb{R}_+^{|\Sigma|}$. Suppose $\nu\in\mathbb{P}(\Sigma)$ is the true probability vector and $\mu\in\mathbb{P}(\Sigma)$ is the nominal fixed probability vector.

A. Finite Alphabet Case: Problem III.1

The total variational distance set is defined by

$$\mathbb{B}_R(\mu)\overset{\triangle}{=}\left\{\nu\in\mathbb{P}(\Sigma):||\nu-\mu||_{TV}\triangleq\sum_{i\in\Sigma}|\nu_i-\mu_i|\leq R\right\}. \quad (19)$$

Let $\ell\triangleq\{\ell_1,\ldots,\ell_{|\Sigma|}\}$ be a non-negative sequence of real numbers, that is $\ell\in\mathbb{R}_+^{|\Sigma|}$. The extremum problem is defined by

$$\mathbb{L}_1(\nu^*)\overset{\triangle}{=}\max_{\nu\in\mathbb{B}_R(\mu)}\sum_{i\in\Sigma}\ell_i\nu_i. \quad (20)$$

Next, we apply the results of Section IV to characterize the optimal ν^* for any $R\in[0,2]$.

By defining $\xi_i\overset{\triangle}{=}\nu_i-\mu_i, i=1,\ldots,|\Sigma|$ and $\xi\in\mathbb{M}_0(\Sigma)$, Problem III.1 can be reformulated as follows.

$$\max_{\nu\in\mathbb{B}_R(\mu)}\sum_{i\in\Sigma}\ell_i\nu_i\longrightarrow\sum_{i\in\Sigma}\ell_i\mu_i+\max_{\xi\in\widetilde{\mathbb{B}}_R(\mu)}\sum_{i\in\Sigma}\ell_i\xi_i. \quad (21)$$

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

$$\sum_{i\in\Sigma}|\xi_i|\leq R, \quad \sum_{i\in\Sigma}\xi_i=0, \quad 0\leq\xi_i+\mu_i\leq 1, \quad \forall i\in\Sigma. \quad (22)$$

By Jordan Decomposition Theorem[5],

$$\sum_{i\in\Sigma}\xi_i=\sum_{i\in\Sigma}\xi_i^+-\sum_{i\in\Sigma}\xi_i^- \quad \text{and} \quad \sum_{i\in\Sigma}|\xi_i|=\sum_{i\in\Sigma}\xi_i^++\sum_{i\in\Sigma}\xi_i^-,$$

and hence, $\sum_{i\in\Sigma}\xi_i^-\equiv\frac{\alpha}{2}\equiv\sum_{i\in\Sigma}\xi_i^+$, where, $\sum_{i\in\Sigma}\xi_i=0$, $\alpha=\sum_{i\in\Sigma}|\xi_i|\leq R$. In addition

$$\sum_{i\in\Sigma}\ell_i\xi_i=\sum_{i\in\Sigma}\ell_i\xi_i^+-\sum_{i\in\Sigma}\ell_i\xi_i^-. \quad (23)$$

Define the maximum and minimum values 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 $(k+1)^{th}$ smallest value by

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

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

till all the elements of Σ are exhausted. Also define the corresponding value of the sequence on these sets by $\ell(\Sigma_k) \triangleq \min_{i\in\Sigma\setminus\Sigma^0\cup(\bigcup_{j=1}^{k}\Sigma_{j-1})} \ell_i$, for all $k \in \{1,2,\ldots,r\}$, where r is the number of Σ_k sets which is at most $|\Sigma\setminus\Sigma^0\cup\Sigma_0|$.
Below we give the solution of Problem III.1.

Theorem V.1 *The maximum pay-off of Problem III.1 is given by*

$$D^+(R) = \ell_{\max}\nu^*(\Sigma^0) + \ell_{\min}\nu^*(\Sigma_0) + \sum_{k=1}^{r}\ell(\Sigma_k)\nu^*(\Sigma_k). \quad (24)$$

Moreover, the optimal probabilities are given by

$$\nu^*(\Sigma^0) \triangleq \sum_{i\in\Sigma^0}\nu_i^* = \sum_{i\in\Sigma^0}\mu_i + \alpha, \quad (25a)$$

$$\nu^*(\Sigma_0) \triangleq \sum_{i\in\Sigma_0}\nu_i^* = \left(\sum_{i\in\Sigma_0}\mu_i - \alpha\right)^+, \quad (25b)$$

$$\nu^*(\Sigma_k) \triangleq \sum_{i\in\Sigma_k}\nu_i^* = \left(\sum_{i\in\Sigma_k}\mu_i - \left(\alpha - \sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i\right)^+\right)^+, \quad (25c)$$

$$\alpha = \min\left(\frac{R}{2}, 1 - \sum_{i\in\Sigma^0}\mu_i\right), \quad (25d)$$

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

Proof. Follows from the next Lemma. ■

Lemma V.2 *(a) If $\sum_{i\in\Sigma_0}\mu_i - \frac{\alpha}{2} \geq 0$ then:*
1. Upper Bound.

$$\sum_{i\in\Sigma}\ell_i\xi_i^+ \leq \ell_{\max}\left(\frac{\alpha}{2}\right). \quad (26)$$

The bound holds with equality if

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

and the optimal probability on Σ^0 is equal to

$$\nu^*(\Sigma^0) \triangleq \sum_{i\in\Sigma^0}\nu_i^* = \min\left(1, \sum_{i\in\Sigma^0}\mu_i + \frac{\alpha}{2}\right). \quad (27)$$

2. Lower Bound.

$$\sum_{i\in\Sigma}\ell_i\xi_i^- \geq \ell_{\min}\left(\frac{\alpha}{2}\right). \quad (28)$$

The bound holds with equality if

$$\sum_{i\in\Sigma_0}\mu_i - \frac{\alpha}{2} \geq 0, \quad \sum_{i\in\Sigma_0}\xi_i^- = \frac{\alpha}{2}, \quad \xi_i^- = 0 \ \text{ for } \ i\in\Sigma\setminus\Sigma_0,$$

and the optimal probability on Σ_0 is equal to

$$\nu^*(\Sigma_0) \triangleq \sum_{i\in\Sigma_0}\nu_i^* = \left(\sum_{i\in\Sigma_0}\mu_i - \frac{\alpha}{2}\right)^+. \quad (29)$$

(b) If $\sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i - \frac{\alpha}{2} \leq 0$ for any $k\in\{1,2,\ldots,r\}$ then:

$$\sum_{i\in\Sigma}\ell_i\xi_i^- \geq \ell(\Sigma_k)\left(\frac{\alpha}{2} - \sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i\right) + \sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\ell_i\mu_i. \quad (30)$$

Moreover, equality holds if

$$\sum_{i\in\Sigma_{j-1}}\xi_i^- = \sum_{i\in\Sigma_{j-1}}\mu_i, \quad \text{for all} \quad j = 1,2,\ldots,k, \quad (31a)$$

$$\sum_{i\in\Sigma_k}\xi_i^- = \left(\frac{\alpha}{2} - \sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i\right), \quad (31b)$$

$$\sum_{j=0}^{k}\sum_{i\in\Sigma_j}\mu_i \geq \frac{\alpha}{2}, \quad (31c)$$

$$\xi_i^- = 0 \quad \text{for all} \quad i\in\Sigma\setminus\Sigma_0\cup\Sigma_1\cup\ldots\cup\Sigma_k, \quad (31d)$$

and the optimal probability on Σ_k sets is given by

$$\nu^*(\Sigma_k) \triangleq \sum_{i\in\Sigma_k}\nu_i^* = \left(\sum_{i\in\Sigma_k}\mu_i - \left(\frac{\alpha}{2} - \sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i\right)^+\right)^+. \quad (32)$$

Proof. See [9]. ■

B. Finite Alphabet Case: Problem III.2

The definitions of the maximum and minimum values of the sequence ℓ and its corresponding support sets are already defined for Problem III.1. 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

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

till all the elements of Σ are exhausted. Also, define the corresponding value of the sequence on these sets by $\ell\left(\Sigma^k\right) \triangleq \max_{i\in\Sigma\setminus\Sigma_0\cup(\bigcup_{j=1}^{k}\Sigma^{j-1})}\ell_i$, for all $k\in\{1,2,\ldots,r\}$, where r is the number of Σ^k sets which is at most $|\Sigma\setminus\Sigma^0\cup\Sigma_0|$.
Below we give the solution of Problem III.2.

Theorem V.3 *The minimum pay-off of Problem III.2 is given by*

$$R^-(D) = \sum_{i\in\Sigma}|\nu_i^* - \mu_i|, \quad (33)$$

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

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

$$R^-(D) = \quad (34)$$
$$\frac{2\left(D - \ell_{\min}\sum_{i\in\Sigma_0}\mu_i - \ell\left(\Sigma^k\right)\sum_{j=1}^{k}\sum_{i\in\Sigma^{j-1}}\mu_i - \sum_{j=k}^{r}\sum_{i\in\Sigma^j}\ell_i\mu_i\right)}{\ell_{\min} - \ell\left(\Sigma^k\right)}.$$

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

$$R^-(D) = \frac{2\left(D - \sum_{i\in\Sigma}\ell_i\mu_i\right)}{\ell_{\min} - \ell_{\max}}. \quad (35)$$

Proof. See [9]. ■

Given Theorem V.1 and Theorem V.3 we can also obtain the solution to the related problems of Section III-A.

VI. EXAMPLES

In this section, we illustrate through examples how the theoretical results obtained in preceding sections are applied.

A. *Working Example* 1

This is a simple yet useful example of Problem III.1. The optimal solution is found by implementing Theorem V.1 for a fixed value of R.

Let $\Sigma = \{i : i = 1,2,3\}$ and for simplicity consider an ascending sequence of lengths $\ell = \{\ell \in \mathbb{R}_+^3 : \ell_1 < \ell_2 < \ell_3\}$ with corresponding nominal probability vector $\mu \in \mathbb{P}_1(\Sigma)$. Specifically, let $\ell = [4,6,8]$ and $\mu = \left[\frac{2}{6}, \frac{2}{6}, \frac{2}{6}\right]$. The sets which correspond to the maximum, minimum and the remaining length are equal to $\Sigma^0 = \{3\}$, $\Sigma_0 = \{1\}$ and $\Sigma_1 = \{2\}$, respectively. Let $R = \frac{1}{3}$. From Theorem V.1, $\alpha = \min\left(\frac{R}{2}, 1-\mu_3\right) = \frac{1}{6}$, and the optimal probabilities are given by

$$\nu^*\left(\Sigma^0\right) = \mu_3 + \alpha = \frac{3}{6}, \qquad \nu^*\left(\Sigma_0\right) = (\mu_1 - \alpha)^+ = \frac{1}{6},$$

$$\nu^*\left(\Sigma_1\right) = \left(\mu_2 - (\alpha - \mu_1)^+\right)^+ = \left(\frac{2}{6} - \left(\frac{1}{6} - \frac{2}{6}\right)^+\right)^+ = \frac{2}{6}.$$

Hence, for $R = \frac{1}{3}$, the maximum pay-off (24), is given by

$$D^+(R) = \ell_3 \nu^*(\Sigma^0) + \ell_1 \nu^*(\Sigma_0) + \ell_2 \nu^*(\Sigma_1) = \frac{40}{6}.$$

B. *Working Example* 2

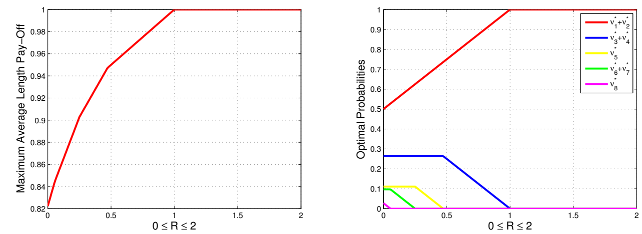
In this example, Problem III.1 is solved for the case in which the sequence $\ell = \{\ell_1, \ldots, \ell_{|\Sigma|}\} \in \mathbb{R}_+^{|\Sigma|}$ consists of a number of lengths which some of them are equal to each other.

Let $\Sigma = \{i : i = 1,2,\ldots,8\}$ and consider a non-ascending 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 $\mu = \left[\frac{23}{72}, \frac{13}{72}, \frac{10}{72} \frac{9}{72}, \frac{8}{72}, \frac{4}{72}, \frac{3}{72}, \frac{2}{72}\right]$. The sets which correspond to the maximum, minimum and all the remaining lengths are equal to $\Sigma^0 = \{1,2\}$, $\Sigma_0 = \{8\}$ and $\Sigma_1 = \{7,6\}$, $\Sigma_2 = \{5\}$, $\Sigma_3 = \{4,3\}$, hence a reduced partition of the space Σ is obtained. Figure 1a depicts $D^+(R)$ given in Theorem V.1, while Figure 1b depicts the optimal probabilities as a function of $R \in [0,2]$. Note that, $D^+(R)$ is a non-decreasing concave function of R and also that is constant in $[R_{\max}, 2]$, where $R_{\max} = 2\left(1 - \mu(\Sigma^0)\right) = 1$.

C. *Working Example* 3

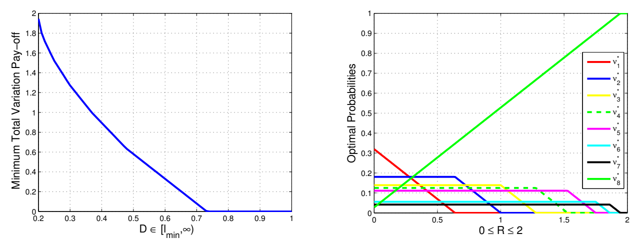
In this example, Problem III.2 is solved for the case in which the sequence $\ell = \{\ell_1, \ldots, \ell_{|\Sigma|}\} \in \mathbb{R}_+^{|\Sigma|}$ consists of a number of lengths which are not equal to each other.

Let $\Sigma = \{i : i = 1,2,\ldots,8\}$ and 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 $\mu = \{\mu \in \mathbb{R}_+^8 : \mu_i > \mu_j, \forall i < j, i,j = 1,\ldots,8\} \in \mathbb{P}_1(\Sigma)$. Specifically, let $\ell = [1,0.8,0.7,0.6,0.5,0.4,0.3,0.2]$, and $\mu = \left[\frac{23}{72}, \frac{13}{72}, \frac{10}{72}, \frac{9}{72}, \frac{8}{72}, \frac{4}{72}, \frac{3}{72}, \frac{2}{72}\right]$. The sets which correspond to the maximum, minimum and all the remaining lengths are equal to $\Sigma^0 = \{1\}$, $\Sigma_0 = \{8\}$ and $\Sigma^1 = \{2\}$, $\Sigma^2 = \{3\}$, $\Sigma^3 = \{4\}$, $\Sigma^4 = \{5\}$, $\Sigma^5 = \{6\}$, $\Sigma^6 = \{7\}$. Figure 2a depicts $R^-(D)$ given in Theorem V.3, while Figure 2b depicts the optimal probabilities as a function of $R \in [0,2]$. $R^-(D)$ is a non-increasing convex function of D, $D \in [\ell_{\min}, \infty)$. Note that for $D < \ell_{\min} = 0.2$ no solution exists and $R^-(D)$ is zero in $[D_{\max}, \infty)$ where $D_{\max} = \sum_{i=1}^8 \ell_i \mu_i = 0.73$. For $D \in [D_{\max}, \infty)$, ν^* is equal to μ and this corresponds to the beginning of Figure 2b.



(a) Maximum pay-off $D^+(R)$. (b) Optimal Probabilities.

Fig. 1: Example B: Maximization of a Linear Functional Pay-Off Subject to Total Variational Distance Constraint.



(a) Minimum Pay-Off $R^-(D)$. (b) Optimal Probabilities.

Fig. 2: Example C: Minimization of Total Variational Distance Subject to a Linear Functional Constraint.

VII. CONCLUSIONS

The paper considered extremum problems with total variational distance metric on the space of measures. Two such problems are formulated, and their solution are discussed for arbitrary spaces, while closed form expressions of the extremum measures are derived for finite alphabet spaces. Examples are presented to show the behavior of the solutions.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley and Sons, Inc., 1991.

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

[3] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. New York: John Wiley & Sons, Inc., 1997.

[4] K. Vastola and H. Poor, "On robust wiener-kolmogorov theory," *IEEE Transactions on Information Theory*, vol. 30, no. 2, pp. 315–327, Mar. 1984.

[5] N. Dunford and J. T. Schwartz, *Linear Operators: Part 1: General Theory*. New York: Interscience Publishers, Inc., 1957.

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

[7] H. Poor, "On robust wiener filtering," *IEEE Transactions on Automatic Control*, vol. 25, no. 3, pp. 531–536, Jun. 1980.

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

[9] C. D. Charalambous, I. Tzortzis, S. Loyka, and T. Charalambous, "Extremum Problems with Total Variation Distance and their Applications," *ArXiv e-prints*, Jan. 2013.