Stochastic optimization in a cumulative prospect theory framework*

Cheng Jie♯, Prashanth L.A.†, Michael Fu$, Steve Marcus‡ and Csaba Szepesvári♮

***Abstract*—Cumulative prospect theory (CPT) is a popular approach for modeling human preferences. It is based on probabilistic distortions and generalizes expected utility theory. We bring CPT to a stochastic optimization framework and propose algorithms for both estimation and optimization of CPT-value objectives. We propose an empirical distribution function-based scheme to estimate the CPT-value and then use this scheme in the inner loop of a CPT-value optimization procedure. We propose both gradient-based as well as gradient-free CPT-value optimization algorithms that are based on two well-known simulation optimization ideas: simultaneous perturbation stochastic approximation (SPSA) and model-based parameter search (MPS), respectively. We provide theoretical convergence guarantees for all the proposed algorithms and also illustrate the potential of CPT-based criteria in a traffic signal control application.**

***Index Terms*—Cumulative prospect theory, stochastic optimization, simultaneous perturbation stochastic approximation, reinforcement learning.**

I. Introduction

In this paper we consider *stochastic optimization problems* where a designer optimizes the system to produce outcomes that are maximally aligned with the preferences of one or possibly multiple humans. As a running example, consider traffic optimization where the goal is to maximize travelers' satisfaction, a challenging problem in big cities. In this example, the outcomes ("return") are travel times, or delays. To capture human preferences, the outcomes are mapped to a single numerical quantity. While preferences of rational agents facing decisions with stochastic outcomes can be modeled using expected utilities, i.e., the expectation of a nonlinear transformation, such as the exponential function, of the rewards or costs [1], [2], humans are subject to various emotional and cognitive biases. As the psychology literature points out, human preferences are inconsistent with expected utilities regardless of what nonlinearities are used [3], [4], [5]. An approach that gained strong support amongst psychologists, behavioral scientists and economists (cf. [6], [7]) is based on Kahneman and Tversky's [5] celebrated *prospect theory* (PT), the theory that we will also base our models of human preferences on in this work. More precisely, we will use *cumulative prospect theory* (CPT), a later, refined variant of prospect theory due to Tversky and Kahneman [8], which superseded prospect theory. CPT generalizes expected utility theory in that in addition to having a utility function transforming the outcomes, another function is introduced which distorts the probabilities in the cumulative distribution function. As compared to prospect theory, CPT is monotone with respect to stochastic dominance, a property that is thought to be useful and more consistent with human preferences.

Our contributions[1]

To the best of our knowledge, we are the first to incorporate CPT into an online stochastic optimization framework. Although on the surface the combination may seem straightforward, in fact there are many challenges that arise from trying to optimize a CPT objective in the stochastic optimization framework, as we will soon see. We outline these challenges as well as our approach to addressing them below.

The first challenge stems from the fact that the CPT-value assigned to a random variable is defined through a nonlinear transformation of the cumulative distribution function associated with the underlying random variable (see Section II for the definitions). Hence, even the problem of estimating the CPT-value given a random sample is challenging. In this paper, we consider a natural quantile-based estimator and analyze its behavior. Under certain technical assumptions, we prove consistency and give sample complexity bounds, the latter based on the Dvoretzky-Kiefer-Wolfowitz (DKW) theorem [10, Chapter 2]. As an example, we show that the sample complexity to estimate the CPT-value for Lipschitz probability distortion weight functions is $O\left(\frac{1}{\epsilon^2}\right)$, for a given accuracy ϵ. This sample complexity coincides with the canonical rate for Monte Carlo-type schemes and is thus unimprovable. Since weight functions that fit well to human preferences are only Hölder continuous (see (A1) in Section III), we also consider this case and find that (unsurprisingly) the sample complexity degrades to $O\left(\frac{1}{\epsilon^{2/\alpha}}\right)$ where $\alpha \in (0, 1]$ is the weight function's Hölder exponent.

* Supported by NSF under Grants CMMI-1434419, CNS-1446665, and CMMI-1362303, by AFOSR under Grant FA9550-15-10050 and the Alberta Innovates Technology Futures through the Alberta Machine Intelligence Institute, NSERC.

♯ Department of Mathematics, University of Maryland, College Park, Maryland, E-Mail: cjie@math.umd.edu,

† Department of Computer Science and Engg., Indian Institute of Technology Madras, Chennai, (work done as a postdoctoral researcher at the Institute for Systems Research, University of Maryland, College Park, Maryland), E-Mail: prashla@cse.iitm.ac.in,

$ Robert H. Smith School of Business & Institute for Systems Research, University of Maryland, College Park, Maryland, E-Mail: mfu@isr.umd.edu,

‡ Department of Electrical and Computer Engineering & Institute for Systems Research, University of Maryland, College Park, Maryland, E-Mail: marcus@umd.edu,

♮ Department of Computing Science, University of Alberta, E-Mail: csaba.szepesvari@ualberta.ca.

[1] A preliminary version of this paper was published in ICML 2016 [9]. In comparison to the conference version, this paper includes additional theoretical results, formal proofs of convergence of both estimation and optimization algorithms, additional experiments and a revised presentation.

Our results on estimating CPT-values form the basis of the algorithms that we propose to maximize CPT-values based on interacting either with a real environment or with a simulator. We consider a smooth parameterization of the CPT-value and propose two algorithms for updating the CPT-value parameter. The first algorithm is a stochastic gradient scheme that uses two-point randomized gradient estimators, borrowed from simultaneous perturbation stochastic approximation (SPSA) [11][2]. The second algorithm is a gradient-free method that is adapted from [12]. The idea is to use a reference model that eventually concentrates on the global minimum and then empirically approximate this reference distribution well-enough. The latter is achieved via natural exponential families in conjunction with Kullback-Leibler (KL) divergence to measure the "distance" from the reference distribution. Guaranteeing convergence of the aforementioned CPT-value optimization algorithms is challenging because only *biased* estimates of the CPT-value are available. We propose a particular way of controlling the arising bias-variance tradeoff and establish convergence for all proposed algorithms.

Related work. Various risk measures have been proposed in the literature, e.g., mean-variance tradeoff [13], exponential utility [14], value at risk (VaR) and conditional value at risk (CVaR) [15]. A large body of literature involves risk-sensitive optimization in the context of Markov decision processes (MDPs). The stochastic optimization context of this paper translates to a risk-sensitive reinforcement learning (RL) problem, and it has been observed in earlier works that risk-sensitive RL is generally hard to solve. For instance, in [16], [17] and [18], the authors provide NP-hardness results for finding a globally variance-optimal policy in discounted and average reward MDPs. Solving CVaR constrained MDPs is equally complicated (cf. [19], [20]). In contrast, incorporating CPT-based criteria incurs extra sample complexity in estimation as compared to that of the classic sample mean estimator for expected value, while the optimization schemes based either on SPSA or model-based parameter search [12] that we propose converge at the same rate as that of their expected value counterparts. In the context of an abstract MDP setting, a CPT-based risk measure has been proposed in [21]. Unlike [21], *(i)* we do not assume a nested structure for the CPT-value, and this implies the lack of a Bellman equation for our CPT measure; *(ii)* we do not assume model information, i.e., we operate in a more general stochastic optimization setting; *(iii)* we develop both estimation and optimization algorithms with convergence guarantees for the CPT-value function. More recently, the authors in [22] incorporate CPT-based criteria into a multi-armed bandit setting, while employing the estimation scheme that we proposed in the shorter version of this paper [9].

The rest of the paper is organized as follows: In Section II, we define the notion of CPT-value for a general random variable. In Section III, we describe the empirical distribution-based scheme for estimating the CPT-value of any random variable. In Section IV, we present gradient-based algorithms for optimizing the CPT-value. We provide proofs of convergence for all the proposed algorithms in Section V. In Sections VI and VII, we present simulation experiments for synthetic and traffic signal control problems, respectively. Finally, in Section VIII we provide our concluding remarks.

[2]A second-order CPT-value optimization scheme based on SPSA is described in [9].

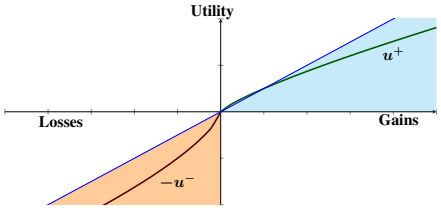


Fig. 1. An example of a utility function. A reference point on the x axis serves as the point of separating gains and losses. For losses, the disutility $-u^-$ is typically convex and for gains, the utility u^+ is typically concave; both functions are non-decreasing and take the value of zero at the reference point.

II. CPT-value

For a real-valued random variable X, we introduce a "CPT-functional" that replaces the traditional expectation operator. The functional, denoted by $\mathbb{C}$, depends on function pairs $u = (u^+, u^-)$ and $w = (w^+, w^-)$. As illustrated in Figure 1, $u^+, u^- : \mathbb{R} \to \mathbb{R}_+$ are continuous, with $u^+(x) = 0$ when $x \leq 0$ and non-decreasing otherwise, and with $u^-(x) = 0$ when $x \geq 0$ and non-increasing otherwise. The functions $w^+, w^- : [0,1] \to [0,1]$, as shown in Figure 2, are continuous, non-decreasing and satisfy $w^+(0) = w^-(0) = 0$ and $w^+(1) = w^-(1) = 1$. The CPT-functional is defined as

$$\mathbb{C}(X) = \int_0^\infty w^+\left(\mathbb{P}\left(u^+(X) > z\right)\right) dz - \int_0^\infty w^-\left(\mathbb{P}\left(u^-(X) > z\right)\right) dz\,. \tag{1}$$

Consider the case when w^+, w^- are identity functions, $u^+(x) = x$ for $x \geq 0$ and 0 otherwise, and $u^-(x) = -x$ for $x \leq 0$ and 0 otherwise. Then, letting $(a)^+ = \max(a, 0)$, $(a)^- = \max(-a, 0)$, we have $\mathbb{C}(X) = \int_0^\infty \mathbb{P}\left(X > z\right) dz - \int_0^\infty \mathbb{P}\left(-X > z\right) dz = \mathbb{E}\left[(X)^+\right] - \mathbb{E}\left[(X)^-\right]$, showing the connection to expectations.

In the definition, u^+, u^- are utility functions corresponding to gains ($X \geq 0$) and losses ($X \leq 0$), respectively, where zero is chosen as an arbitrary "reference point" to separate gains and losses. Handling losses and gains separately is a salient feature of CPT, and this addresses the tendency of humans to play safe with gains and take risks with losses. To illustrate this tendency, consider a scenario where one can either earn \$500 with probability (w.p.) 1 or earn \$1000 w.p. 0.5 and nothing otherwise. The human tendency is to choose the former option of a certain gain. If we flip the situation, i.e., a certain loss of \$500 or a loss of \$1000 w.p. 0.5, then humans choose the latter option. This distinction of playing safe with gains and taking risks with losses is captured by a concave gain-utility u^+ and a convex disutility $-u^-$, as illustrated in Fig. 1.

The functions w^+, w^-, called the weight functions, capture the idea that humans deflate high-probabilities and inflate low-probabilities. For example, humans usually choose a stock that

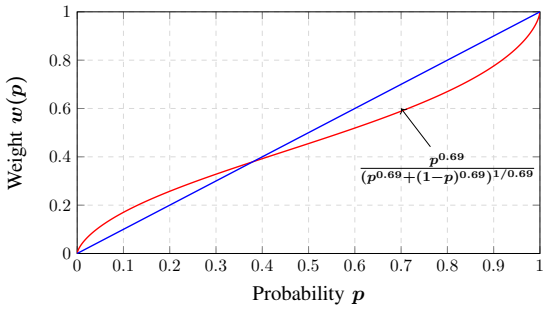


Fig. 2. An example of a weight function. A typical CPT weight function inflates small, and deflates large probabilities, capturing the tendency of humans doing the same when faced with decisions of uncertain outcomes.

gives a large reward, e.g., one million dollars w.p. $1/10^6$, over one that gives \$1 w.p. 1 and the reverse when signs are flipped. Thus the value seen by a human subject is nonlinear in the underlying probabilities – an observation backed by strong empirical evidence [8], [23]. In [8], the authors recommend $w(p) = \frac{p^\eta}{(p^\eta+(1-p)^\eta)^{1/\eta}}$, while in [24], the author recommends $w(p) = \exp(-(-\ln p)^\eta)$, with $0 < \eta < 1$. In both cases, the weight function has an inverted-s shape.

Remark 1. *(RL application) For any RL problem setting, one can define the return for a given policy and then apply a CPT-functional on the return. For instance, with a fixed policy, the random variable (r.v.) X could be the total reward in a stochastic shortest path problem or the infinite horizon cumulative reward in a discounted MDP.*

Remark 2. *(Generalization) As noted earlier, the CPT-value is a generalization of mathematical expectation. It is also possible to obtain* (1) *to coincide with other risk measures, e.g. value at risk (VaR) and conditional value at risk (CVaR), by appropriate choice of weight functions.*

III. CPT-value estimation

We devise a scheme for estimating the CPT-value $\mathbb{C}(X)$ given only samples from the distribution of X. Before diving into the details of CPT-value estimation, let us discuss the conditions necessary for the CPT-value to be well-defined. Observe that the first integral in (1), i.e., $\int_0^{+\infty} w^+\left(\mathbb{P}\left(u^+(X) > z\right)\right) dz$ may diverge even if the first moment of random variable $u^+(X)$ is finite. For example, suppose U has the tail distribution function $\mathbb{P}\left(U > z\right) = \frac{1}{z^2}, z \in [1, +\infty)$, and $w^+(z)$ takes the form $w(z) = z^{\frac{1}{3}}$. Then, the first integral in (1), i.e., $\int_1^{+\infty} z^{-\frac{2}{3}}\, dz$ does not even exist. A similar argument applies to the second integral in (1).

To overcome the integrability issues, we assume that the weight functions w^+, w^- satisfy one of the following assumptions for continuous valued r.v.s:

Assumption (A1). The weight functions $w^\pm$ are Hölder continuous with common order α and constant H, i.e., $\sup_{x\neq y} \frac{|w^\pm(x)-w^\pm(y)|}{|x-y|^\alpha} \leq H,\ \forall x, y \in [0,1]$. Further, there exists $\gamma \leq \alpha$ such that (s.t.) $\int_0^{+\infty} \mathbb{P}^\gamma(u^+(X) > z)dz < +\infty$ and $\int_0^{+\infty} \mathbb{P}^\gamma(u^-(X) > z)dz < +\infty$, where $\mathbb{P}^\gamma(\cdot) = (\mathbb{P}(\cdot))^\gamma$.

Assumption (A1'). The weight functions w^+, w^- are Lipschitz with common constant L, and $u^+(X)$ and $u^-(X)$ both have bounded first moments.

Proposition 1. *Under (A1) or (A1'), the CPT-value $\mathbb{C}(X)$ as defined by* (1) *is finite.*

Proof. See Section V-A1. □

(A1'), even though it implies (A1), is a useful special case because it does away with additional assumptions required to establish asymptotic consistency under (A1). For the theoretical results, we also require the following assumption on the utility functions:

Assumption (A2). The utility functions u^+ and $-u^-$ are continuous and non-decreasing on their support $\mathbb{R}^+$ and $\mathbb{R}^-$, respectively.

Finally, we also analyze the setting where X is a discrete valued r.v. Such a setting is common in practice and carries the additional advantage that, under a local Lipschitz assumption on the distribution of X, one gets better sample complexity as compared to those under (A1) and (A1').

A. CPT-value estimation using quantiles

Let ξ_k^+ and ξ_k^- denote the kth quantiles of the r.v.s $u^+(X)$ and $u^-(X)$, respectively. Then, it can be seen that (see Proposition 2 in Section V-A1)

$$\lim_{n\to\infty} \sum_{i=1}^{n} \xi^+_{\frac{i}{n}} \left(w^+\left(\frac{n+1-i}{n}\right) - w^+\left(\frac{n-i}{n}\right)\right) = \int_0^{+\infty} w^+\left(\mathbb{P}\left(u^+(X) > z\right)\right) dz. \tag{2}$$

A similar claim holds with $u^-(X), \xi_k^-, w^-$ in place of $u^+(X)$, ξ_α^+, w^+, respectively.

However, we do not know the distribution of $u^+(X)$ or $u^-(X)$ and hence, we next present a procedure that uses order statistics for estimating quantiles and this in turn assists estimation of the CPT-value along the lines of (2). The estimation scheme is presented in Algorithm 1.

Algorithm 1 CPT-value estimation

1: **Input:** samples $X_1, \ldots, X_n$ from the distribution of X.
2: Arrange the samples in ascending order and label them as follows: $X_{[1]}, X_{[2]}, \ldots, X_{[n]}$.
3: Let

$$\overline{\mathbb{C}}_n^+ := \sum_{i=1}^{n} u^+(X_{[i]}) \left(w^+\left(\frac{n+1-i}{n}\right) - w^+\left(\frac{n-i}{n}\right)\right),$$

$$\overline{\mathbb{C}}_n^- := \sum_{i=1}^{n} u^-(X_{[i]}) \left(w^-\left(\frac{i}{n}\right) - w^-\left(\frac{i-1}{n}\right)\right).$$

4: Return $\overline{\mathbb{C}}_n = \overline{\mathbb{C}}_n^+ - \overline{\mathbb{C}}_n^-$.

Consider the special case when $w^+(p) = w^-(p) = p$ and u^+ ($-u^-$), when restricted to the positive (respectively, negative) half line, are the identity functions. In this case, the CPT-value estimator $\overline{\mathbb{C}}_n$ coincides with the sample mean estimator for regular expectation.

Notice that the CPT estimator $\overline{\mathbb{C}}_n$ in Algorithm 1 can be written equivalently as follows:

$$\overline{\mathbb{C}}_n = \int_0^\infty w^+\left(1-\hat{F}_n^+(x)\right)dx - \int_0^\infty w^-\left(1-\hat{F}_n^-(x)\right)dx. \tag{3}$$

The above relation holds because

$$\begin{aligned}
&\sum_{i=1}^n u^+\left(X_{[i]}\right)\left(w^+\left(\frac{n+1-i}{n}\right)-w^+\left(\frac{n-i}{n}\right)\right)\\
&=\sum_{i=1}^{n-1} w^+\left(\frac{n-i}{n}\right)\left(u^+\left(X_{[i+1]}\right)-u^+\left(X_{[i]}\right)\right)+u^+(X_{[1]})\\
&=\int_0^\infty w^+\left(1-\hat{F}_n^+(x)\right)dx, \text{ and}\\
&\sum_{i=1}^n u^-\left(X_{[i]}\right)\left(w^-\left(\frac{i}{n}\right)-w^-\left(\frac{i-1}{n}\right)\right)\\
&=\int_0^\infty w^-\left(1-\hat{F}_n^-(x)\right)dx,
\end{aligned}$$

where $\hat{F}_n^+(x)$ and $\hat{F}_n^-(x)$ are the empirical distributions of $u^+(X)$ and $u^-(X)$, respectively.

B. Results for Hölder and Lipschitz continuous weights

Proposition 2. ***(Asymptotic consistency)*** *Assume (A1), (A2), $F^+(\cdot)$ and $F^-(\cdot)$, the respective distribution functions of $u^+(X)$ and $u^-(X)$, are Lipschitz continuous on the respective intervals $(0,+\infty)$, and $(-\infty,0)$, and the utility functions u^+, u^- satisfy*

$$\lim_{n\to\infty}\frac{u^+(X_{[n]})}{n^\alpha}\to 0 \text{ and } \lim_{n\to\infty}\frac{u^-(X_{[n]})}{n^\alpha}\to 0 \text{ a.s.},$$

where α is the Hölder exponent for $w^\pm$.

Then, we have

$$\overline{\mathbb{C}}_n \to \mathbb{C}(X) \text{ a.s. as } n\to\infty \tag{4}$$

where $\overline{\mathbb{C}}_n$ is as defined in Algorithm 1 and $\mathbb{C}(X)$ as in (1)*.*

The conditions on utility functions above are satisfied by popular distribution choices such as Gaussian and exponential, but not by heavy-tailed distributions, e.g. Cauchy.

Proof. See Section V-A1. □

Under an additional assumption on the utility functions, our next result shows that $O\left(\frac{1}{\epsilon^{2/\alpha}}\right)$ number of samples are sufficient to get a high-probability estimate of the CPT-value that is ϵ-accurate.

Proposition 3. ***(Sample complexity.)*** *Assume (A1), (A2) and also that the utilities $u^+(X)$ and $u^-(X)$ are bounded by a constant M. Then, $\forall\epsilon>0$, we have*

$$\mathbb{P}\left(\left|\overline{\mathbb{C}}_n-\mathbb{C}(X)\right|\geq\epsilon\right)\leq 2e^{-2n\left(\frac{\epsilon}{HM}\right)^{\frac{2}{\alpha}}} \tag{5}$$

Instead, if the utilities functions are sub-Gaussian[3]*, then $\forall\epsilon>0$ and $n\geq\left(\frac{\ln 2-\ln\epsilon}{2\alpha}\right)^{\alpha+2}$, we have*

$$\mathbb{P}\left(\left|\overline{\mathbb{C}}_n-\mathbb{C}(X)\right|\geq\epsilon\right)\leq 2ne^{-n^{\frac{1}{2+\alpha}}}+2e^{-n^{\frac{1}{2+\alpha}}\left(\frac{\epsilon}{2H}\right)^{\frac{2}{\alpha}}} \tag{6}$$

[3]A r.v. X with mean μ is sub-Gaussian if $\exists\sigma>0$ such that $\mathbb{E}\left[e^{\lambda(X-\mu)}\right]\leq e^{\sigma^2\lambda^2/2}, \forall\lambda\in\mathbb{R}$.

Corollary 1. *Assume (A1), (A2). If utilities $u^+(X)$ and $u^-(X)$ are bounded by M, then*

$$\mathbb{E}\left|\overline{\mathbb{C}}_n-\mathbb{C}(X)\right|\leq\frac{(8HM)\,\Gamma(\alpha/2)}{n^{\alpha/2}}.$$

Instead, if the utilities are sub-Gaussian, then

$$\mathbb{E}\left|\overline{\mathbb{C}}_n-\mathbb{C}(X)\right|\leq\frac{\Gamma(\frac{1}{2})\cdot\alpha\cdot 2^{1-\alpha}}{n^{\frac{\alpha}{\alpha+2}}}+\frac{\Gamma(\frac{1}{2})\cdot\sqrt{2}(2H)^{\frac{2}{\alpha}}}{n^{\frac{2-\alpha}{2+\alpha}}}$$

Proof. See Section V-A1. □

Setting $\alpha=1$, one can obtain the asymptotic consistency claim in Proposition 2 for Lipschitz weight functions. However, this result is under a restrictive Lipschitz assumption on the distribution functions of $u^+(X)$ and $u^-(X)$. Using a different proof technique and (A1') in place of (A1), we can obtain a result similar to Proposition 2 without a Lipschitz assumption on the distribution functions. The following claim makes this precise.

Proposition 4. ***(Asymptotic consistency)*** *Assume (A1') and (A2). Then, we have*

$$\overline{\mathbb{C}}_n\to\mathbb{C}(X) \text{ a.s. as } n\to\infty.$$

Proof. See Section V-A2. □

Setting $\alpha=1$ in Proposition 3, we observe that one can achieve the canonical Monte Carlo rate for Lipschitz continuous weights. Choosing the weights to be the identity function, we observe that the sample complexity cannot be improved. On the other hand, for Hölder continuous weights, we incur a sample complexity of order $O\left(\frac{1}{\epsilon^{2/\alpha}}\right)$ for accuracy $\epsilon>0$ and this is generally worse than the canonical Monte Carlo rate of $O\left(\frac{1}{\epsilon^2}\right)$, for $\alpha<1$. An interesting question here is if the sample complexity from Proposition 3 be improved upon, say to $O(1/\epsilon^2)$ for achieving ϵ accuracy? The next result shows that the best achievable sample complexity, in the minimax sense, is $\Omega\left(\frac{1}{\epsilon^{2/\alpha}}\right)$ over the class of Hölder-continuous weight functions.

Before presenting the lower bound, we define the notion of minimax error. Let $\mathcal{P}$ be a nonempty set of distributions. Let $\mathbb{C}(P)$ denote the CPT-value of a r.v. with distribution $P\in\mathcal{P}$ and $\overline{C}_n:\mathbb{R}^n\to\mathbb{R}$ denote an estimator. The minimax error $\mathcal{R}_n(\mathcal{P})$ is defined by

$$\mathcal{R}_n(\mathcal{P}):=\inf_{\overline{C}_n}\sup_{P\in\mathcal{P}}\mathbb{E}_{X_{1:n}\sim P^{\otimes n}}\left|\overline{C}_n(X_{1:n})-\mathbb{C}(P)\right| \tag{7}$$

Proposition 5. ***(Lower bound)*** *For a set of distributions $\mathcal{P}$ supported within the interval $[0,1]$, the minimax error satisfies*

$$\mathcal{R}_n(\mathcal{P})\geq\frac{1}{4(6n)^{\frac{\alpha}{2}}}, \text{ for all } n\geq 1.$$

Proof. See Section V-A3. □

C. Locally Lipschitz weights and discrete-valued X

Here we assume that X is a discrete valued r.v. with finite support. Let $p_i, i=1,\ldots,K$, denote the probability

of incurring a gain/loss $x_i, i = 1, \ldots, K$, where $x_1 \leq \ldots \leq x_l \leq 0 \leq x_{l+1} \leq \ldots \leq x_K$ and let

$$F_k = \sum_{i=1}^{k} p_i \text{ if } k \leq l \text{ and } \sum_{i=k}^{K} p_i \text{ if } k > l. \tag{8}$$

In this setting, the first integral, say $\mathbb{C}^+(X)$, in the definition of CPT-value (1) can be simplified as follows:

$$\begin{aligned}
\mathbb{C}^+(X) &= \int_0^{u^+(x_{l+1})} w^+\left(\mathbb{P}\left(u^+(X) > z\right)\right) dz \\
&\quad + \sum_{k=l+1}^{K-1} \int_{u^+(x_k)}^{u^+(x_{k+1})} w^+\left(\mathbb{P}\left(u^+(X) > z\right)\right) dz \\
&\quad + \int_{u^+(x_K)}^{\infty} w^+\left(\mathbb{P}\left(u^+(X) > z\right)\right) dz \\
&= w^+(F_{l+1})u^+(x_{l+1}) + \sum_{i=l+2}^{K} w^+(F_i)(u^+(x_i) - u^+(x_{i-1})) \\
&= \sum_{i=l+1}^{K-1} u^+(x_i)\Big(w^+(F_i) - w^+(F_{i+1})\Big) + u^+(x_K)w^+(p_K).
\end{aligned}$$

The second integral in (1) can be simplified in a similar fashion, and we obtain the following form for the overall CPT-value of a discrete-valued X:

$$\begin{aligned}
\mathbb{C}(X) = &\Big(\sum_{i=l+1}^{K-1} u^+(x_i)\Big(w^+(F_i) - w^+(F_{i+1})\Big) + u^+(x_K)w^+(p_K)\Big) \\
&- \Big((u^-(x_1))w^-(p_1) + \sum_{i=2}^{l} u^-(x_i)\Big(w^-(F_i) - w^-(F_{i-1})\Big)\Big).
\end{aligned}$$

Estimation scheme: Let $X_1, \ldots, X_n$ be n samples from the distribution of X. Define $\hat{p}_k := \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i = x_k\}}$ and

$$\hat{F}_k = \sum_{i=1}^{k} \hat{p}_k \text{ if } k \leq l \text{ and } \sum_{i=k}^{K} \hat{p}_k \text{ if } k > l. \tag{9}$$

Then, we estimate $\mathbb{C}(X)$ as follows:

$$\begin{aligned}
\overline{\mathbb{C}}_n = &\Big(\sum_{i=l+1}^{K-1} u^+(x_i)\Big(w^+(\hat{F}_i) - w^+(\hat{F}_{i+1})\Big) + u^+(x_K)w^+(\hat{p}_K)\Big) \\
&- \Big(u^-(x_1)w^-(\hat{p}_1) + \sum_{i=2}^{l} u^-(x_i)\Big(w^-(\hat{F}_i) - w^-(\hat{F}_{i-1})\Big)\Big).
\end{aligned}$$

Assumption (A1"). The weight functions w^+ and w^- are locally Lipschitz continuous, i.e., for any $k = 1, \ldots, K$, there exist $L_k < \infty$ and $\rho_k > 0$, such that, for $k = 1, \ldots, l$,

$$|w^-(F_k) - w^-(p)| \leq L_k|F_k - p|, \quad \forall p \in (F_k - \rho_k, F_k + \rho_k),$$

and for $k = 1 + 1, \ldots, K$,

$$|w^+(F_k) - w^+(p)| \leq L_k|F_k - p|, \quad \forall p \in (F_k - \rho_k, F_k + \rho_k).$$

Proposition 6. *Assume (A1"). Let $L = \max_{k=1,\ldots,K} L_k$ and $\rho = \min\{\rho_k\}$, where L_k and ρ_k are as defined in (A1"). Let $M = \max\{u^-(x_k), k = 1, \ldots, l\} \bigcup \{u^+(x_k), k = l + 1, \ldots, K\}$. Then, $\forall \epsilon > 0, \delta > 0$, we have*

$$\mathbb{P}\left(\left|\overline{\mathbb{C}}_n - \mathbb{C}(X)\right| \leq \epsilon\right) > 1 - \delta, \forall n \geq \frac{1}{\kappa}\ln\left(\frac{1}{\delta}\right)\ln\left(\frac{4K}{M}\right),$$

where $\kappa = \min(\rho^2, \epsilon^2/(KLM)^2)$.

In comparison to Propositions 3 and 4, observe that the sample complexity for discrete X scales with the local Lipschitz constant L, and this can be much smaller than the global Lipschitz constant of the weight functions, or the weight functions may not be Lipschitz globally.

Proof. See Section V-A4. □

A variant of Corollary 1 can be obtained by integrating the high-probability bound in Proposition 6; we omit the details here.

IV. CPT-value optimization

A. Optimization objective

Suppose the r.v. X in (1) is a function of a d-dimensional parameter θ. In this section we consider the problem

$$\text{Find } \theta^* = \arg\max_{\theta \in \Theta} \mathbb{C}(X^\theta), \tag{10}$$

where Θ is a compact and convex subset of $\mathbb{R}^d$. The above problem encompasses policy optimization in an MDP that can be discounted or average or stochastic shortest path and/or partially observed. The difference here is that we apply the CPT-functional to the return of a policy, instead of using the expected return.

B. Gradient algorithm using SPSA (CPT-SPSA)

Gradient estimation: Given that we operate in a learning setting and only have asymptotically unbiased estimates of the CPT-value from Algorithm 1, we require a simulation scheme to estimate $\nabla\mathbb{C}(X^\theta)$. Simultaneous perturbation methods are a general class of stochastic gradient schemes that optimize a function given only noisy sample values - see [25] for a textbook introduction. SPSA is a well-known scheme that estimates the gradient using two sample values. In our context, at any iteration n of CPT-SPSA, with parameter θ_n, the gradient $\nabla\mathbb{C}(X^{\theta_n})$ is estimated as follows: For any $i = 1, \ldots, d$,

$$\widehat{\nabla}_i\mathbb{C}(X^\theta) = \frac{\overline{\mathbb{C}}_n^{\theta_n + \delta_n\Delta_n} - \overline{\mathbb{C}}_n^{\theta_n - \delta_n\Delta_n}}{2\delta_n\Delta_n^i}, \tag{11}$$

where δ_n is a positive scalar that satisfies (A3) below, $\Delta_n = \left(\Delta_n^1, \ldots, \Delta_n^d\right)^\dagger$, where $\{\Delta_n^i, i = 1, \ldots, d\}$, $n = 1, 2, \ldots$ are i.i.d. symmetric, ± 1-valued Bernoulli r.v.s, independent of $\theta_0, \ldots, \theta_n$ and $\overline{\mathbb{C}}_n^{\theta_n + \delta_n\Delta_n}$ (resp. $\overline{\mathbb{C}}_n^{\theta_n - \delta_n\Delta_n}$) denotes the CPT-value estimate that uses m_n samples of the r.v. $X^{\theta_n + \delta_n\Delta_n}$ (resp. $\overline{X}^{\theta_n - \delta_n\Delta_n}$). The (asymptotic) unbiasedness of the gradient estimate is proven in Lemma 4.

Update rule: We incrementally update the parameter θ in the ascent direction as follows:

$$\theta_{n+1} = \Pi\left(\theta_n + \gamma_n\widehat{\nabla}\mathbb{C}(X^{\theta_n})\right), \tag{12}$$

where γ_n is a step-size chosen to satisfy (A3) below and $\Pi = (\Pi_1, \ldots, \Pi_d)$ is an operator that ensures that the update (12) stays bounded within the compact and convex set Θ.

On the number of samples m_n per iteration: Recall that the CPT-value estimation scheme is asymptotically unbiased, i.e., providing samples with parameter θ_n at instant n, we obtain its CPT-value estimate as $\mathbb{C}(X^{\theta_n}) + \psi_n^\theta$, with ψ_n^θ denoting the error in estimation. The estimation error can be controlled by increasing the number of samples m_n in each iteration of CPT-SPSA. This is unlike many simulation optimization settings where one only sees function evaluations with zero mean noise and there is no question of deciding on m_n to control the estimation error as we have in our setting.

To motivate the choice for m_n, we first rewrite the update rule (12) as follows:

$$\theta_{n+1}^i = \Pi_i\left(\theta_n^i + \gamma_n\Big(\frac{\mathbb{C}(X^{\theta_n+\delta_n\Delta_n}) - \mathbb{C}(X^{\theta_n-\delta_n\Delta_n})}{2\delta_n\Delta_n^i}\Big) + \kappa_n\right),$$

where $\kappa_n = \frac{(\psi_n^{\theta_n+\delta_n\Delta_n} - \psi_n^{\theta_n-\delta_n\Delta_n})}{2\delta_n\Delta_n^i}$. Let $\zeta_n = \sum_{l=0}^n \gamma_l\kappa_l$. Then, a critical requirement that allows us to ignore the estimation error term ζ_n is the following condition (see Lemma 1 in Chapter 2 of [26]):

$$\sup_{l\geq 0}(\zeta_{n+l} - \zeta_n) \to 0 \text{ as } n \to \infty.$$

While Theorems 2–3 show that the estimation error ψ^θ is bounded above, to establish convergence of the CPT-SPSA, we increase the number of samples m_n so that the bias vanishes asymptotically. The assumption below provides a condition on the increase rate of m_n.

Assumption (A3). The step-sizes γ_n and the perturbation constants δ_n are positive $\forall n$ and satisfy

$$\gamma_n, \delta_n \to 0, \frac{1}{m_n^{\alpha/2}\delta_n} \to 0, \sum_n \gamma_n = \infty \text{ and } \sum_n \frac{\gamma_n^2}{\delta_n^2} < \infty.$$

While the conditions on γ_n and δ_n are standard for SPSA-based algorithms, the condition on m_n is motivated by the earlier discussion. A simple choice that satisfies the above conditions is $\gamma_n = a_0/n$, $m_n = m_0 n^\nu$ and $\delta_n = \delta_0/n^\gamma$, for some $\nu, \gamma > 0$ with $\gamma > \nu\alpha/2$.

Assumption (A4). The CPT-value $\mathbb{C}(X^\theta)$ is a continuously differentiable function of θ, with bounded third derivative.

In a typical RL setting involving finite state action spaces, a sufficient condition for ensuring (A4) holds is to assume that the policy is continuously differentiable in θ.

Convergence result for CPT-SPSA

We use the ordinary differential equation (ODE) method for establishing asymptotic convergence of CPT-SPSA. Consider the ODE:

$$\dot{\theta}_t^i = \check{\Pi}_i\left(-\nabla\mathbb{C}(X^{\theta_t^i})\right), \text{ for } i = 1, \ldots, d, \tag{13}$$

where $\check{\Pi}_i(f(\theta)) := \lim_{\vartheta\downarrow 0}\frac{\Pi_i(\theta+\vartheta f(\theta))-\theta}{\vartheta}$, for any continuous $f(\cdot)$. Let $\mathcal{K} \subset \{\theta^* \mid \check{\Pi}_i\left(\nabla_i\mathbb{C}(X^{\theta^*})\right) = 0, \forall i = 1, \ldots, d\}$ denote the set of asymptotically stable equilibrium points of the ODE (13). That $\mathcal{K} \neq \phi$ can be inferred by using the fact that $\mathbb{C}(X^\theta)$ itself serves as a Lyapunov function for (13) (see Section V-B1 for details).

Theorem 1. *Assume (A1)-(A4). Then, $\mathcal{K} \neq \phi$ and for θ_n governed by* (12)*, we have*

$$\theta_n \to \mathcal{K} \text{ a.s. as } n \to \infty.$$

Proof. See Section V-B1. □

Let $\mathcal{K}' = \{\theta^* \mid \nabla_i\mathbb{C}(X^{\theta^*}) = 0, \forall i = 1, \ldots, d\}$ denote the set of critical points of the CPT-value. If $\mathcal{K}'$ lies within the set Θ onto which the iterate θ_n (updated according to (12)) is projected, then the above theorem ensures that CPT-SPSA converges to $\mathcal{K}'$. When it not possible to ensure that $\mathcal{K}' \subset \Theta$, the iterate θ_n might get stuck on the boundary of Θ.

Remark 3. *The convergence result presented for CPT-SPSA is applicable to more general settings where an algorithm is provided samples of a performance objective, with an estimation error that vanishes asymptotically. Examples of such settings are average reward optimization via policy gradient methods in an RL context [27] or in the context of an optimal stopping problem [28].*

C. Model-based parameter search algorithm (CPT-MPS)

In this section, we provide a gradient-free algorithm (CPT-MPS) for maximizing the CPT-value, that is based on the MRAS$_2$ algorithm proposed by Chang e al. [12]. While CPT-SPSA is a local optimization scheme, CPT-MPS converges to the global optimum, say θ^*, for the problem (10), assuming one exists.

To illustrate the main idea in the algorithm, assume we know the form of $\mathbb{C}(X^\theta)$. Then, the idea is to generate a sequence of reference distributions $g_k(\theta)$ on the parameter space Θ, such that it eventually concentrates on the global optimum θ^*. One simple way, suggested in Chapter 4 of [12] is

$$g_k(\theta) = \frac{\mathcal{H}(\mathbb{C}(X^\theta))g_{k-1}(\theta)}{\int_\Theta \mathcal{H}(\mathbb{C}(X^{\theta'}))g_{k-1}(\theta')\nu(d\theta')}, \quad \forall\, \theta \in \Theta, \tag{14}$$

where ν is the Lebesgue/counting measure on Θ and $\mathcal{H}$ is a strictly decreasing function. The above construction for g_k's assigns more weight to parameters having higher CPT-values.

Next, consider a setting where one can obtain the CPT-value $\mathbb{C}(X^\theta)$ (without any noise) for any parameter θ. In this case, we consider a family of parameterized distributions, say $\{f(\cdot,\eta), \eta \in \mathbb{C}\}$ and incrementally update the distribution parameter η such that it minimizes the following KL divergence: $\mathcal{D}(g_k, f(\cdot,\eta)) := \int_\Theta \ln\frac{g_k(\theta)}{f(\theta,\eta)} g_k(\theta)\nu(d\theta)$, where $\hat{\theta}$ is a random vector taking values in the parameter space Θ. As recommended in [12], we employ the natural exponential family (NEF) for the family of distributions $f(\cdot,\theta)$, since it ensures that the KL distance above can be computed analytically. An algorithm to optimize CPT-value in this *noiseless* setting would perform the following update:

$$\eta_{n+1} \in \arg\max_{\eta\in\mathbb{C}} E_{\eta_n}\left[\frac{[\mathcal{H}(\mathbb{C}(X^{\hat{\theta}})]^n}{f(\hat{\theta},\eta_n)} \ln f(\hat{\theta},\eta)\right], \tag{15}$$

where $E_{\eta_n}[\mathbb{C}(X^{\hat{\theta}})] = \int_\Theta \mathbb{C}(X^\theta) f(\theta,\eta_n)\nu(d\theta)$.

Algorithm 2 presents the pseudocode for the CPT-value optimization setting where we obtain only asymptotically

Algorithm 2 Structure of CPT-MPS algorithm.

Input: family of distributions $\{f(\cdot,\eta)\}$, initial parameter vector η_0 s.t. $f(\theta,\eta_0)>0\,\forall\,\theta\in\Theta$, trajectory lengths $\{m_n\}$, $\rho_0\in(0,1]$, $N_0>1$, $\varepsilon>0$, $\varsigma>1$, $\lambda\in(0,1)$, strictly increasing function $\mathcal{H}$and $\chi_{-1}=-\infty$.
for $n=0,1,2,\ldots$ **do**
Generate N_n parameters $\Lambda_n=\{\theta_n^1,\ldots,\theta_n^{N_n}\}$ using the mixture distribution $\widetilde{f}(\cdot,\eta_n)=(1-\lambda)f(\cdot,\widetilde{\eta}_n)+\lambda f(\cdot,\eta_0)$.
for $i=1,2,\ldots,N_n$ **do**
Obtain CPT-value estimate $\overline{\mathbb{C}}_n^{\theta_n^i}$ using m_n samples.
end for
Elite Sampling:
Order the CPT-value estimates as $\{\overline{\mathbb{C}}_n^{\theta_n^{(1)}},\ldots,\overline{\mathbb{C}}_n^{\theta_n^{(N_n)}}\}$.
Compute the $(1-\rho_n)$-quantile $\widetilde{\chi}_n(\rho_n,N_n)=\overline{\mathbb{C}}_n^{\theta_n^{\lceil(1-\rho_n)N_n\rceil}}$.
Thresholding:
find largest $\bar{\rho}\in(0,\rho_n)$ such that $\widetilde{\chi}_n(\bar{\rho},N_n)\geq\bar{\chi}_{n-1}+\varepsilon$;
if $\bar{\rho}$ exists **then**
Set $\bar{\chi}_n=\widetilde{\chi}_n(\bar{\rho},N_n)$, $\rho_{n+1}=\bar{\rho}$, $N_{n+1}=N_n$, $\theta_n^*=\theta_{1-\bar{\rho}}$.
else
Set $\bar{\chi}_n=\overline{\mathbb{C}}_n^{\theta_{n-1}^*}$, $\rho_{n+1}=\rho_n$, $N_{n+1}=\lceil\varsigma N_n\rceil$, $\theta_n^*=\theta_{n-1}^*$.
end if
Sampling distribution update:

$$\eta_{n+1}\in\arg\max_{\eta\in\mathbb{C}}\sum_{i=1}^{N_n}\frac{[\mathcal{H}(\overline{\mathbb{C}}^{\theta_n^i})]^n)}{\widetilde{f}(\theta,\eta_n)}\widetilde{I}\big(\overline{\mathbb{C}}^{\theta_n^i},\bar{\chi}_n\big)\ln f(\theta,\eta),$$

where $\widetilde{I}(z,\chi):=0$ if $z\leq\chi-\varepsilon$, $(z-\chi+\varepsilon)/\varepsilon$ if $\chi-\varepsilon<z<\chi$ and 1 if $z\geq\chi$.
end for
Return θ_n

unbiased estimates of the CPT-value $\mathbb{C}(X^\theta)$ for any parameter θ. As in [12], we use only an elite portion of the candidate parameters that have been sampled, as this guides the parameter search procedure towards better regions more efficiently in comparison to an alternative that uses all the candidate parameters for updating η.

The main convergence result is stated below.

Theorem 2. *Assume (A1), (A2) and that $m_n\to\infty$ as $n\to\infty$. Suppose that multivariate normal densities are used for the sampling distribution, i.e., $\eta_n=(\mu_n,\Sigma_n)$, where μ_n and Σ_n denote the mean and covariance of the normal densities. Then,*

$$\lim_{n\to\infty}\mu_n=\theta^*\ \textit{and}\ \lim_{n\to\infty}\Sigma_n=0_{d\times d}\quad a.s. \tag{16}$$

Proof. See Section V-B2. □

V. Convergence Proofs

A. Proofs for CPT-value estimator

1) Hölder continuous weights: For proving Propositions 2 and 6, we require Hoeffding's inequality, which is given below.

Lemma 1. *Let $Y_1,\ldots Y_n$ be independent random variables satisfying $\mathbb{P}(a\leq Y_i\leq b)=1$, $\forall i$, where $a<b$. Then, $\forall\epsilon>0$,*

$$\mathbb{P}\left(\left|\sum_{i=1}^n Y_i-\sum_{i=1}^n E(Y_i)\right|\geq n\epsilon\right)\leq 2\exp\{-2n\epsilon^2/(b-a)^2\}.$$

Proof. (***Proposition 1***)
Hölder continuity of w^+ and $w^+(0)=0$ imply that

$$\int_0^\infty w^+\left(\mathbb{P}\left(u^+(X)>z\right)\right)dz\leq H\int_0^\infty\mathbb{P}^\alpha\left(u^+(X)>z\right)dz$$
$$\leq H\int_0^\infty\mathbb{P}^\gamma\left(u^+(X)>z\right)dz<\infty.$$

The second inequality is valid since $\mathbb{P}(u^+(X)>z)\leq 1$. The claim follows for the first integral in (1), and the finiteness of the second integral in (1) can be argued in an analogous fashion. □

We now state and prove a lemma that will be used in the proof of Proposition 2.

Lemma 2. *Assume (A1). Let $\xi^+_{\frac{i}{n}}$ and $\xi^-_{\frac{i}{n}}$ denote the $\frac{i}{n}$th quantile of $u^+(X)$ and $u^-(X)$, respectively. Then, we have*

$$\lim_{n\to\infty}\sum_{i=1}^{n-1}\xi^+_{\frac{i}{n}}\left(w^+\left(\frac{n-i}{n}\right)-w^+\left(\frac{n-i-1}{n}\right)\right)$$
$$=\int_0^\infty w^+\left(\mathbb{P}\left(u^+(X)>z\right)\right)dz<\infty, \tag{17}$$
$$\lim_{n\to\infty}\sum_{i=1}^{n-1}\xi^-_{\frac{i}{n}}\left(w^-\left(\frac{i}{n}\right)-w^-\left(\frac{i-1}{n}\right)\right)$$
$$=\int_0^\infty w^-\left(\mathbb{P}\left(u^-(X)>z\right)\right)dz<\infty. \tag{18}$$

Proof. We will focus on proving equation (17). For all $z\in(0,+\infty)$, the following convergence claim holds w.p.1:

$$\sum_{i=1}^{n-1}w^+\left(\frac{i}{n}\right)I_{\left[\xi^+_{\frac{n-i-1}{n}},\xi^+_{\frac{n-i}{n}}\right]}(z)\xrightarrow{n\to\infty}w^+\left(\mathbb{P}\left(u^+(X)>z\right)\right). \tag{19}$$

To infer the above claim, observe that since $u^+(X)$ ranges in $(0,+\infty),\forall z$, there exists i such that $z\in[\xi^+_{\frac{n-i-1}{n}},\xi^+_{\frac{n-i}{n}}]$, which implies that

$$w^+\left(\mathbb{P}\left(u^+(X)\geq z\right)\right)\in\left[w^+\left(\frac{i}{n}\right),w^+\left(\frac{i+1}{n}\right)\right].$$

Hence, we have

$$\left|\sum_{j=1}^{n-1}w^+\left(\frac{j}{n}\right)I_{\left[\xi^+_{\frac{n-j-1}{n}},\xi^+_{\frac{n-j}{n}}\right]}(z)-w^+\left(\mathbb{P}\left(u^+(X)>z\right)\right)\right|$$
$$\leq\left|w^+\left(\frac{i}{n}\right)-w^+\left(\frac{i+1}{n}\right)\right|$$

Since w^+ is Hölder continuous, we have

$$\left|w^+\left(\frac{i}{n}\right)-w^+\left(\frac{i+1}{n}\right)\right|\xrightarrow{n\to\infty}0,$$

and the claim in (19) follows.

Further, for all $z\in[0,\infty)$,

$$\sum_{j=1}^{n-1}w^+\left(\frac{j}{n}\right)I_{\left[\xi^+_{\frac{n-j-1}{n}},\xi^+_{\frac{n-j}{n}}\right]}(z)<w^+\left(\mathbb{P}\left(u^+(X)>z\right)\right). \tag{20}$$

The integral of the LHS of (19) can be simplified as follows:

$$\begin{aligned}
&\int_0^\infty \sum_{j=0}^{n} w^+\left(\frac{j}{n}\right) I_{\left[\xi^+_{\frac{n-j-1}{n}},\xi^+_{\frac{n-j}{n}}\right]}(z)dz \\
&= \sum_{j=0}^{n-1} w^+\left(\frac{j}{n}\right)\left(\xi^+_{\frac{n-j}{n}} - \xi^+_{\frac{n-j-1}{n}}\right) \\
&= \sum_{j=0}^{n-1} \xi^+_{\frac{j}{n}}\left(w^+\left(\frac{n-j}{n}\right) - w^+\left(\frac{n-j-1}{n}\right)\right). \qquad (21)
\end{aligned}$$

Now, the main claim in (17) can be inferred from (19),(20) and (21) in conjunction with the dominated convergence theorem.

The second part of (17) follows in a similar fashion. □

Proof. (***Proposition 2***)
Without loss of generality, assume that w^+ and w^- are both $(1,\alpha)$ Hölder. We prove the claim for the first integral in the CPT-value estimator $\overline{\mathbb{C}}_n$ in Algorithm 1, i.e., we show that

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{i=1}^{n} u^+\left(X_{[i]}\right)\left(w^+\left(\frac{n-i+1}{n}\right) - w^+\left(\frac{n-i}{n}\right)\right) \\
&= \int_0^\infty w^+\left(P\left(u^+(X) > z\right)\right) dz, \text{ a.s.} \qquad (22)
\end{aligned}$$

The main part of the proof is focused on finding an upper bound for the probability

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_{i=1}^{n-1} u^+\left(X_{[i]}\right)\left(w^+\left(\frac{n-i}{n}\right) - w^+\left(\frac{n-i-1}{n}\right)\right)\right.\right. \\
&\left.\left. - \sum_{i=1}^{n-1} \xi^+_{\frac{i}{n}}\left(w^+\left(\frac{n-i}{n}\right) - w^+\left(\frac{n-i-1}{n}\right)\right)\right| > \epsilon\right).
\end{aligned}$$

Observe the fact that

$$\begin{aligned}
&\sum_{i=1}^{n} u^+\left(X_{[i]}\right)\left(w^+\left(\frac{n-i+1}{n}\right) - w^+\left(\frac{n-i}{n}\right)\right) \\
&- \sum_{i=1}^{n-1} u^+\left(X_{[i]}\right)\left(w^+\left(\frac{n-i}{n}\right) - w^+\left(\frac{n-i-1}{n}\right)\right) \\
&= \sum_{i=1}^{n} \left(u^+\left(X_{[i]}\right) - u^+\left(X_{[i-1]}\right)\right) w^+\left(\frac{n+1-i}{n}\right) \\
&- \sum_{i=1}^{n} \left(u^+\left(X_{[i]}\right) - u^+\left(X_{[i-1]}\right)\right) w^+\left(\frac{n-i}{n}\right) \\
&= \sum_{i=1}^{n} \left(u^+\left(X_{[i]}\right) - u^+\left(X_{[i-1]}\right)\right) \\
&\qquad \times \left(w^+\left(\frac{n+1-i}{n}\right) - w^+\left(\frac{n-i}{n}\right)\right) \\
&\leq u^+\left(X_{[n]}\right) \times \frac{1}{n^\alpha}
\end{aligned}$$

Under (A1), the term $\frac{u^+\left(X_{[n]}\right)}{n^\alpha}$ converges to 0. Hence, for the asymptotic convergence of estimator, thanks to Lemma 2, it suffices to show that

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\left(\left|\sum_{i=1}^{n-1} u^+\left(X_{[i]}\right)\left(w^+\left(\frac{n-i}{n}\right) - w^+\left(\frac{n-i-1}{n}\right)\right)\right.\right. \\
&\left.\left. - \sum_{i=1}^{n-1} \xi^+_{\frac{i}{n}}\left(w^+\left(\frac{n-i}{n}\right) - w^+\left(\frac{n-i-1}{n}\right)\right)\right| > \epsilon\right) = 0.
\end{aligned}$$

For any given $\epsilon > 0$, we have

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_{i=1}^{n-1} u^+\left(X_{[i]}\right)\left(w^+\left(\frac{n-i}{n}\right) - w^+\left(\frac{n-i-1}{n}\right)\right)\right.\right. \\
&\qquad \left.\left. - \sum_{i=1}^{n-1} \xi^+_{\frac{i}{n}}\left(w^+\left(\frac{n-i}{n}\right) - w^+\left(\frac{n-i-1}{n}\right)\right)\right| > \epsilon\right) \\
&\leq \mathbb{P}\left(\bigcup_{i=1}^{n-1}\left\{\left|u^+\left(X_{[i]}\right)\left(w^+\left(\frac{n-i}{n}\right) - w^+\left(\frac{n-i-1}{n}\right)\right)\right.\right.\right. \\
&\qquad \left.\left.\left. - \xi^+_{\frac{i}{n}}\left(w^+\left(\frac{n-i}{n}\right) - w^+\left(\frac{n-i-1}{n}\right)\right)\right| > \frac{\epsilon}{n-1}\right\}\right) \\
&\leq \sum_{i=1}^{n-1} \mathbb{P}\left(\left|\left(u^+\left(X_{[i]}\right) - \xi^+_{\frac{i}{n}}\right)\right.\right. \\
&\qquad \left.\left. \times \left(w^+\left(\frac{n-i}{n}\right) - w^+\left(\frac{n-i-1}{n}\right)\right)\right| > \frac{\epsilon}{n-1}\right) \\
&\leq \sum_{i=1}^{n-1} \mathbb{P}\left(\left|\left(u^+\left(X_{[i]}\right) - \xi^+_{\frac{i}{n}}\right)\left(\frac{1}{n}\right)^\alpha\right| > \frac{\epsilon}{n-1}\right) \qquad (23) \\
&\leq \sum_{i=1}^{n-1} \mathbb{P}\left(\left|\left(u^+\left(X_{[i]}\right) - \xi^+_{\frac{i}{n}}\right)\right| > \frac{\epsilon}{n^{1-\alpha}}\right). \qquad (24)
\end{aligned}$$

In the above, (23) follows from the fact that w^+ is Hölder with constant 1.

Now we find an upper bound for the probability of a single term in the sum above, i.e.,

$$\begin{aligned}
&\mathbb{P}\left(\left|u^+\left(X_{[i]}\right) - \xi^+_{\frac{i}{n}}\right| > \frac{\epsilon}{n^{(1-\alpha)}}\right) = \mathbb{P}\left(u^+\left(X_{[i]}\right) - \xi^+_{\frac{i}{n}} > \frac{\epsilon}{n^{(1-\alpha)}}\right) \\
&+ \mathbb{P}\left(u^+\left(X_{[i]}\right) - \xi^+_{\frac{i}{n}} < -\frac{\epsilon}{n^{(1-\alpha)}}\right).
\end{aligned}$$

We focus on the first term above.

Let $W_j = I_{\left(u^+(X_j) > \xi^+_{\frac{i}{n}} + \frac{\epsilon}{n^{(1-\alpha)}}\right)}, j = 1,\ldots,n.$

Using the fact that a probability distribution function is non-decreasing, we obtain

$$\begin{aligned}
&\mathbb{P}\left(u^+(X_{[i]}) - \xi^+_{\frac{i}{n}} > \frac{\epsilon}{n^{(1-\alpha)}}\right) = \mathbb{P}\left(\sum_{j=1}^{n} W_j > n - i\right) \\
&= \mathbb{P}\left(\sum_{j=1}^{n} W_j > n\left(1 - \frac{i}{n}\right)\right) \\
&= \mathbb{P}\left(\sum_{j=1}^{n} W_j - n\left[1 - F^+\left(\xi^+_{\frac{i}{n}} + \frac{\epsilon}{n^{(1-\alpha)}}\right)\right]\right. \\
&\qquad \left. > n\left[F^+\left(\xi^+_{\frac{i}{n}} + \frac{\epsilon}{n^{(1-\alpha)}}\right) - \frac{i}{n}\right]\right).
\end{aligned}$$

Using the fact that $EW_j = 1 - F^+\left(\xi^+_{\frac{i}{n}} + \frac{\epsilon}{n^{(1-\alpha)}}\right)$ in conjunction with Hoeffding's inequality, we obtain

$$\mathbb{P}\left(\sum_{i=1}^{n} W_j - n\left[1 - F^+\left(\xi^+_{\frac{i}{n}} + \frac{\epsilon}{n^{(1-\alpha)}}\right)\right]\right.$$

$$> n\left[F^+\left(\xi^+_{\frac{i}{n}}+\frac{\epsilon}{n^{(1-\alpha)}}\right)-\frac{i}{n}\right]\Bigg)\leq e^{-2n\delta_i^{'}},$$

where $\delta_i^{'}=F^+\left(\xi^+_{\frac{i}{n}}+\frac{\epsilon}{n^{(1-\alpha)}}\right)-\frac{i}{n}$. Since F^+ is Lipschitz, we have that $\delta_i^{'}\leq L^+\left(\frac{\epsilon}{n^{(1-\alpha)}}\right)$. Hence, we obtain

$$\mathbb{P}\left(u^+(X_{[i]})-\xi^+_{\frac{i}{n}}>\frac{\epsilon}{n^{(1-\alpha)}}\right)\leq e^{-2nL^+\frac{\epsilon}{n^{(1-\alpha)}}}$$
$$=e^{-2n^\alpha L^+\epsilon}. \quad (25)$$

In a similar fashion, one can show that

$$\mathbb{P}\left(u^+(X_{[i]})-\xi^+_{\frac{i}{n}}<-\frac{\epsilon}{n^{(1-\alpha)}}\right)\leq e^{-2n^\alpha L^+\epsilon}. \quad (26)$$

Combining (25) and (26), we obtain

$$\mathbb{P}\left(\left|u^+(X_{[i]})-\xi^+_{\frac{i}{n}}\right|<-\frac{\epsilon}{n^{(1-\alpha)}}\right)\leq 2e^{-2n^\alpha L^+\epsilon}.$$

Plugging the above in (24), we obtain

$$\mathbb{P}\left(\left|\sum_{i=1}^{n-1}u^+\left(X_{[i]}\right)\left(w^+\left(\frac{n-i}{n}\right)-w^+\left(\frac{n-i-1}{n}\right)\right)\right.\right.$$
$$\left.\left.-\sum_{i=1}^{n-1}\xi^+_{\frac{i}{n}}\left(w^+\left(\frac{n-i}{n}\right)-w^+\left(\frac{n-i-1}{n}\right)\right)\right|>\epsilon\right)$$
$$\leq 2(n-1)e^{-2n^\alpha L^+\epsilon}\leq 2ne^{-2n^\alpha L^+\epsilon}. \quad (27)$$

Notice that $\sum_{n=1}^{\infty}2ne^{-2n^\alpha L^+\epsilon}<\infty$ since the sequence $2ne^{-2n^\alpha L^+}$ will decrease faster than the sequence $\frac{1}{n^k}$ provided $k>1$.

By applying the Borel-Cantelli lemma, $\forall\epsilon>0$, we have

$$\mathbb{P}\left(\left|\sum_{i=1}^{n-1}u^+\left(X_{[i]}\right)\left(w^+\left(\frac{n-i}{n}\right)-w^+\left(\frac{n-i-1}{n}\right)\right)\right.\right.$$
$$\left.\left.-\sum_{i=1}^{n-1}\xi^+_{\frac{i}{n}}\left(w^+\left(\frac{n-i}{n}\right)-w^+\left(\frac{n-i-1}{n}\right)\right)\right|>\epsilon, i.o.\right)$$
$$=0,$$

which implies (22).

The proof of $\mathbb{C}_n^-\to\mathbb{C}^-(X)$ follows in a similar manner as above by replacing $u^+(X_{[i]})$ by $u^-(X_{[n-i]})$, after observing that u^- is decreasing, which in turn implies that $u^-(X_{[n-i]})$ is an estimate of the quantile $\xi^-_{\frac{i}{n}}$. □

For proving Proposition 3, we require the DKW inequality, which we recall below.

Lemma 3. ***(DKW inequality)***
Let F denote the cdf of r.v. U and $\hat{F}_n(u)=\frac{1}{n}\sum_{i=1}^n I_{[U_i\leq u]}$ denote the empirical distribution of U, with $U_1,\ldots,U_n$ sampled from F. Then, for any $\epsilon>0$, we have

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}}|\hat{F}_n(x)-F(x)|>\epsilon\right)\leq 2e^{-2n\epsilon^2}.$$

Proof. (***Proposition 3***)
To prove (6), we only need to address the w^+ part, and the w^- part follows in a similar fashion. Observe that for all $c>0$, we have

$$\mathbb{P}\left(|\mathbb{C}_n-\mathbb{C}(X)|>\epsilon\right)\leq\mathbb{P}\left(u^+\left(X_{[i]}\right)\geq n^c\right)$$
$$+\mathbb{P}\left(\{|\mathbb{C}_n-\mathbb{C}(X)|>\epsilon\}\bigcap\left\{u^+\left(X_{[i]}\right)<n^c\right\}\right).$$

On the event $\{u^+\left(X_{[i]}\right)<n^c\}$, we have

$$\left|\int_0^\infty w^+\left(\mathbb{P}\left(u^+(X)>t\right)\right)dt-\int_0^\infty w^+\left(1-\hat{F}_n^+(t)\right)dt\right|$$
$$=\left|\int_0^\infty w^+\left(\mathbb{P}\left(u^+(X)>t\right)\right)dt-\int_0^{n^c} w^+\left(1-\hat{F}_n^+(t)\right)dt\right|.$$

Notice that

$$\int_{n^c}^\infty w\left(\mathbb{P}\left(X>s\right)\right)ds\leq\frac{1}{n^c}\int_{n^c}^\infty\frac{s}{n^c}e^{-2\alpha s^2}ds$$
$$=n^c\frac{4}{\alpha}e^{-2\alpha(n^c)^2}\leq\frac{\epsilon}{2}\text{ for }n\geq\left(\frac{\ln 2-\ln\epsilon}{2\alpha}\right)^{\frac{1}{c}}.$$

Thus, we obtain

$$\mathbb{P}\left(\{|\mathbb{C}_n-\mathbb{C}(X)|>\epsilon\}\bigcap\left\{u^+\left(X_{[n]}\right)<n^c\right\}\right)\leq$$
$$\mathbb{P}\left(\left|\int_0^{n^c}w^+\left(\mathbb{P}\left(u^+(X)>t\right)\right)dt-\int_0^{n^c}w^+\left(1-\hat{F}_n^+(t)\right)dt\right|>\frac{\epsilon}{2}\right).$$

Now, plugging in the DKW inequality, we have

$$\mathbb{P}\left(\left|\int_0^\infty w^+\left(\mathbb{P}\left(u^+(X)>t\right)\right)dt-\int_0^\infty w^+\left(1-\hat{F}_n^+(t)\right)dt\right|>\frac{\epsilon}{2}\right)$$
$$\leq\mathbb{P}\left(Hn^c\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(u^+\left(X\right)<t\right)-\hat{F}_n^+\left(t\right)\right|^\alpha>\frac{\epsilon}{2}\right)$$
$$\leq 2e^{-2n\left(\frac{\epsilon}{2Hn^c}\right)^{\frac{2}{\alpha}}}=2e^{-2n^{1-\frac{2c}{\alpha}}\left(\frac{\epsilon}{2H}\right)^{\frac{2}{\alpha}}}. \quad (28)$$

Meanwhile, by sub-Gaussianity, we infer that

$$\mathbb{P}\left(u^+\left(X_{[n]}\right)>n^c\right)=1-\mathbb{P}\left(u^+\left(X_{[n]}\right)\leq n^c\right)$$
$$=1-\left(\mathbb{P}\left(X_i\leq n^c\right)\right)^n\leq 1-\left(1-e^{-2n^c}\right)^n$$
$$\leq 1-\left(1-ne^{-2n^c}\right)=ne^{-2n^c},$$

where the last inequality is obtained by Taylor approximation. As a result,

$$\mathbb{P}\left(|\mathbb{C}_n-\mathbb{C}(X)|>\epsilon\right)\leq ne^{-2n^c}+2e^{-2n^{1-\frac{2c}{\alpha}}\left(\frac{\epsilon}{2H}\right)^{\frac{2}{\alpha}}}.$$

The right side of the above inequality will be optimized with $c=1-\frac{2c}{\alpha}$, i.e., for $c=\frac{1}{2+\alpha}$. The claim in (6) follows.

To prove (5) under the condition that utilities functions are bounded by M, notice that

$$\left|\int_0^\infty w^+\left(\mathbb{P}\left(u^+(X)>t\right)\right)dt-\int_0^\infty w^+\left(1-\hat{F}_n^+(t)\right)dt\right|$$
$$=\left|\int_0^M w^+\left(\mathbb{P}\left(u^+(X)>t\right)\right)dt-\int_0^M w^+\left(1-\hat{F}_n^+(t)\right)dt\right|$$
$$\leq HM\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(u^+(X)<t\right)-\hat{F}_n^+(t)\right|^\alpha.$$

The bound in (5) can be inferred by replacing n^c by M and $\frac{\epsilon}{2}$ by ϵ in inequality (28). □

Proof. (***Corollary 1***)
When the utilities are bounded by M, integrating the high-probability bound (5) in Proposition 3, we obtain

$$\mathbb{E}\left|\overline{\mathbb{C}}_n-\mathbb{C}(X)\right|\leq\int_0^\infty\mathbb{P}\left(\left|\overline{\mathbb{C}}_n-\mathbb{C}(X)\right|\geq\epsilon\right)d\epsilon$$

$$\leq 4\int_0^\infty \exp\left(-2n\left(\epsilon/HM\right)^{2/\alpha}\right)d\epsilon \leq \frac{8HM\Gamma\left(\alpha/2\right)}{n^{\alpha/2}}. \tag{29}$$

For the sub-Gaussian case, notice that if we truncate $u^+\left(X_{[i]}\right)$ by $n^c\sqrt{\epsilon}$ instead of n^c and repeat the steps used in the proof of Proposition 3, we obtain

$$\mathbb{P}\left(\left|\overline{\mathbb{C}}_n - \mathbb{C}(X)\right| \geq \epsilon\right) \leq ne^{-2n^{2c}\epsilon^{\frac{1}{2}}} + 2e^{-2n^{1-\frac{2c}{\alpha}}\left(\frac{\epsilon^{\frac{1}{2}}}{2H}\right)^{\frac{2}{\alpha}}}.$$

Setting $c = \frac{1}{2+\alpha}$, we obtain the following:

$$\mathbb{E}\left|\overline{\mathbb{C}}_n - \mathbb{C}(X)\right| \leq \frac{\Gamma\left(\frac{1}{2}\right)\cdot\alpha\cdot 2^{1-\alpha}}{n^{\frac{\alpha}{\alpha+2}}} + \frac{\Gamma\left(\frac{1}{2}\right)\sqrt{2}\left(2H\right)^{\frac{2}{\alpha}}}{n^{\frac{2-\alpha}{2+\alpha}}}.$$

□

2) Lipschitz continuous weights: Setting $\alpha = \gamma = 1$ in the proof of Proposition 4, it is easy to see that the CPT-value (1) is finite. We provide a proof of the asymptotic convergence claim in Proposition 4 below.

Proof. (***Proposition 4***)
We first prove the asymptotic convergence claim for the first integral (3) in the CPT-value estimator in Algorithm 1, i.e., we show

$$\int_0^\infty w^+\left(1-\hat{F}_n^+\left(x\right)\right)dx \to \int_0^\infty w^+\left(\mathbb{P}\left(u^+\left(X\right)>x\right)\right)dx. \tag{30}$$

Since w^+ is Lipschitz continuous with, say, constant L, we have almost surely that $w^+\left(1-\hat{F}_n\left(x\right)\right) \leq L\left(1-\hat{F}_n\left(x\right)\right)$, for all n and $w^+\left(\mathbb{P}\left(u^+\left(X\right)>x\right)\right) \leq L\left(\mathbb{P}\left(u^+\left(X\right)>x\right)\right)$, since $w^+\left(0\right)=0$.

We have

$$\int_0^\infty \left(\mathbb{P}\left(u^+\left(X\right)>x\right)\right)dx = \mathbb{E}\left[u^+\left(X\right)\right], \text{ and}$$
$$\int_0^\infty \left(1-\hat{F}_n^+\left(x\right)\right)dx = \int_0^\infty\int_x^\infty d\hat{F}_n\left(t\right)dx. \tag{31}$$

Since $\hat{F}_n^+\left(x\right)$ has bounded support on $\mathbb{R}$ $\forall n$, the integral in (31) is finite. Applying Fubini's theorem to the RHS of (31), we obtain

$$\int_0^\infty\int_x^\infty d\hat{F}_n\left(t\right)dx = \int_0^\infty t d\hat{F}_n\left(t\right) = \frac{1}{n}\sum_{i=1}^n u^+\left(X_{[i]}\right),$$

where $u^+\left(X_{[i]}\right), i=1,\ldots,n$ are the order statistics, i.e., $u^+\left(X_{[1]}\right) \leq \ldots \leq u^+\left(X_{[n]}\right)$.

Notice that

$$\frac{1}{n}\sum_{i=1}^n u^+\left(X_{[i]}\right) = \frac{1}{n}\sum_{i=1}^n u^+\left(X_i\right) \xrightarrow{a.s} \mathbb{E}\left[u^+\left(X\right)\right].$$

From the foregoing,

$$\lim_{n\to\infty}\int_0^\infty L\left(1-\hat{F}_n\left(x\right)\right)dx \xrightarrow{a.s} \int_0^\infty L\left(\mathbb{P}\left(u^+\left(X\right)>x\right)\right)dx.$$

The claim in (30) now follows by invoking the generalized dominated convergence theorem by setting $f_n = w^+(1-\hat{F}_n^+(x))$ and $g_n = L(1-\hat{F}_n(x))$, and noticing that $L(1-\hat{F}_n(x)) \xrightarrow{a.s.} L(\mathbb{P}\left(u^+(X)>x\right))$ uniformly over x. The latter fact is implied by the Glivenko-Cantelli theorem (cf. Chapter 2 of [10]).

Following similar arguments, it is easy to show that

$$\int_0^\infty w^-\left(1-\hat{F}_n^-(x)\right)dx \to \int_0^\infty w^-\left(\mathbb{P}\left(u^-(X)>x\right)\right)dx.$$

The final claim regarding the almost sure convergence of $\overline{\mathbb{C}}_n$ to $\mathbb{C}(X)$ now follows. □

3) Lower bound for estimation error:

Proof. (***Proposition 5***)
We use Le Cam's method to establish the lower bound. Let X_v, $v \in \{-1,+1\}$ denote a Bernoulli r.v. with underlying distribution P_v, $v\in\{+1,-1\}$ defined by

$$P_v(X=1) = \frac{1+v\delta^{\frac{1}{\alpha}}}{2} \text{ and } P_v(X=0) = \frac{1-v\delta^{\frac{1}{\alpha}}}{2},$$

where $\delta \in [0, 2^{-\alpha}]$ is left to be chosen later. Setting $u^+(x) = x, x\geq 0$, $w^+ = w^- = w$, where w is Hölder continuous with exponent $\alpha\in(0,1)$, we have

$$\mathbb{C}(P_v) = w(1+v\delta^{\frac{1}{\alpha}}), \quad v\in\{+1,-1\}.$$

Suppose that w also satisfies the following condition: $|w(p)-w(\tilde{p})| \geq |p-\tilde{p}|^\alpha$ for $p,\tilde{p}\in(0,1)$. An example of such a w for $\alpha = 1/2, H=1$, as suggested in the proof of Theorem 6 in [22], is: $w(p) = \frac{1}{2} - \frac{1}{\sqrt{2}}\sqrt{\frac{1}{2}-p}$ for $p\in[0,1/2]$, and $w(p) = \frac{1}{2} + \frac{1}{\sqrt{2}}\sqrt{p-\frac{1}{2}}$ for $p\in(1/2,1]$.

Setting $p = 1+\delta^{\frac{1}{\alpha}}$ and $\tilde{p} = 1-\delta^{\frac{1}{\alpha}}$, we have

$$|\mathbb{C}(P_{+1}) - \mathbb{C}(P_{-1})| = |w(p)-w(\tilde{p})| \geq |p-\tilde{p}|^\alpha = \delta.$$

By Le Cam's method [29], the minimax error then satisfies

$$\begin{aligned}\mathcal{R}_n(\mathcal{P}) &\geq \frac{\delta}{2}\left(1-\left\|P_{+1}^n - P_{-1}^n\right\|_{\mathrm{TV}}\right)\\ &\geq \frac{\delta}{2}\left(1-\left(\tfrac{1}{2}D_{\mathrm{kl}}\left(P_{+1}^n\|P_{-1}^n\right)\right)^{\frac{1}{2}}\right),\end{aligned} \tag{32}$$

where $P_v^n := \otimes^n P_v$ is the joint distribution of n samples from P_v, $\|\|_{\mathrm{TV}}$ is the total variation distance and (32) follows from Pinsker's inequality. We bound the KL-divergences as follows:

$$\begin{aligned}&D_{\mathrm{kl}}\left(P_-^n\|P_+^n\right) = nD_{\mathrm{kl}}\left(P_+\|P_-\right)\\ &= \frac{n}{2}\left((1-\delta^{\frac{1}{\alpha}})\log\frac{1-\delta^{\frac{1}{\alpha}}}{1+\delta^{\frac{1}{\alpha}}} + (1+\delta^{\frac{1}{\alpha}})\log\frac{1+\delta^{\frac{1}{\alpha}}}{1-\delta^{\frac{1}{\alpha}}}\right)\\ &= n\delta^{\frac{1}{\alpha}}\log\frac{1+\delta^{\frac{1}{\alpha}}}{1-\delta^{\frac{1}{\alpha}}} \leq 3n\delta^{\frac{2}{\alpha}},\end{aligned}$$

where the first equality uses chain rule of KL-divergences, the second follows by the definition of KL-divergences between Bernoullis, and the final inequality follows by using the fact that for $x\in[0,1/2]$, $x\log\frac{1+x}{1-x} \leq 3x^2$.

Plugging the bound on KL-divergences into (32), we obtain

$$\mathcal{R}_n(\mathcal{P}) \geq \frac{\delta}{2}\left(1-\sqrt{\frac{3n}{2}}\delta^{\frac{1}{\alpha}}\right) = \frac{1}{4(6n)^{\frac{\alpha}{2}}}, \tag{33}$$

for $\delta = \frac{1}{(6n)^{\frac{\alpha}{2}}}$. Noting that $\delta\in[0,2^{-\alpha}]$ for any $n\geq 1$ finishes the proof. □

4) Proofs for discrete valued X: Without loss of generality, assume $w^+ = w^- = w$.

Proposition 7. *Let F_k and $\hat{F}_k$ be as defined in* (8)*,* (9)*, Then, for every $\epsilon > 0$,*

$$P(|\hat{F}_k - F_k| > \epsilon) \leq 2e^{-2n\epsilon^2}.$$

Proof. We focus on the case when $k > l$, while the case of $k \leq l$ is proved in a similar fashion.

$$\begin{aligned}
&\mathbb{P}\left(\left|\hat{F}_k - F_k\right| > \epsilon\right) \\
&= \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \geq x_k\}} - \frac{1}{n}\sum_{i=1}^{n} E(I_{\{X_i \geq x_k\}})\right| > \epsilon\right) \\
&= \mathbb{P}\left(\left|\sum_{i=1}^{n} I_{\{X_i \geq x_k\}} - \sum_{i=1}^{n} E(I_{\{X_i \geq x_k\}})\right| > n\epsilon\right) \qquad (34) \\
&\leq 2e^{-2n\epsilon^2}, \qquad (35)
\end{aligned}$$

where the last inequality above follows by an application of Hoeffding inequality after observing that X_i are independent of each other and for each i, the corresponding r.v. in (34) is an indicator that is trivially bounded above by 1. □

Proposition 8. *Under the conditions of Proposition 6, we have*

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_{i=1}^{K} u_k w(\hat{F}_k) - \sum_{i=1}^{K} u_k w(F_k)\right| > \epsilon\right) \\
&\leq K\left(e^{-2n\rho^2} + e^{-2n\epsilon^2/(KLM)^2}\right), \text{ where} \\
&u_k = u^-(x_k) \text{ if } k \leq l \text{ and } u^+(x_k) \text{ if } k > l. \qquad (36)
\end{aligned}$$

Proof. Observe that

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_{k=1}^{K} u_k w(\hat{F}_k) - \sum_{k=1}^{K} u_k w(F_k)\right| > \epsilon\right) \\
&\leq \mathbb{P}\left(\bigcup_{k=1}^{K}\left|u_k w(\hat{F}_k) - u_k w(F_k)\right| > \frac{\epsilon}{K}\right) \\
&\leq \sum_{k=1}^{K}\mathbb{P}\left(\left|u_k w(\hat{F}_k) - u_k w(F_k)\right| > \frac{\epsilon}{K}\right). \qquad (37)
\end{aligned}$$

For each $k = 1, ...K$, the function w is locally Lipschitz on $[p_k - \rho, p_k + \rho)$ with common constant L. Therefore, for each k, we can decompose the corresponding probability in (37) as follows:

$$\begin{aligned}
&\mathbb{P}\left(\left|u_k w(\hat{F}_k) - u_k w(F_k)\right| > \frac{\epsilon}{K}\right) \\
&= \mathbb{P}\left(\left\{\left|F_k - \hat{F}_k\right| > \rho\right\} \bigcap \left\{\left|u_k w(\hat{F}_k) - u_k w(F_k)\right| > \frac{\epsilon}{K}\right\}\right) \\
&+ \mathbb{P}\left(\left\{\left|F_k - \hat{F}_k\right| \leq \rho\right\} \bigcap \left\{\left|u_k w(\hat{F}_k) - u_k w(F_k)\right| > \frac{\epsilon}{K}\right\}\right) \\
&\leq \mathbb{P}\left(\left|F_k - \hat{F}_k\right| > \rho\right) \\
&+ \mathbb{P}\left(\left\{\left|F_k - \hat{F}_k\right| \leq \rho\right\} \bigcap \left\{\left|u_k w(\hat{F}_k) - u_k w(F_k)\right| > \frac{\epsilon}{K}\right\}\right). \qquad (38)
\end{aligned}$$

Using the fact that w is L-Lipschitz together with Proposition 7, we obtain

$$\begin{aligned}
&\mathbb{P}\left(\left\{\left|F_k - \hat{F}_k\right| \leq \rho\right\} \bigcap \left\{\left|u_k w(\hat{F}_k) - u_k w(F_k)\right| > \frac{\epsilon}{K}\right\}\right) \\
&\leq \mathbb{P}\left(u_k L\left|F_k - \hat{F}_k\right| > \frac{\epsilon}{K}\right) \\
&\leq e^{-2n\epsilon/(KLu_k)^2} \leq e^{-2n\epsilon/(KLM)^2}, \forall k. \qquad (39)
\end{aligned}$$

Using Proposition 7, we obtain

$$\mathbb{P}\left(\left|F_k - \hat{F}_k\right| > \rho\right) \leq e^{-2n\rho^2}, \forall k. \qquad (40)$$

Using (39) and (40) in (38), we obtain

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_{k=1}^{K} u_k w(\hat{F}_k) - \sum_{k=1}^{K} u_k w(F_k)\right| > \epsilon\right) \\
&\leq \sum_{k=1}^{K}\mathbb{P}\left(\left|u_k w(\hat{F}_k) - u_k w(F_k)\right| > \frac{\epsilon}{K}\right) \\
&\leq K\left(e^{-2n\rho^2} + e^{-2n\epsilon^2/(KLM)^2}\right).
\end{aligned}$$

The claim follows. □

Proof. (***Proposition 6***)
With u_k as defined in (36), we need to prove that, $\forall n \geq \frac{1}{\kappa}\ln\left(\frac{1}{\delta}\right)\ln\left(\frac{4K}{M}\right)$, the following high-probability bound holds

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_{i=1}^{K} u_k\left(w\left(\hat{F}_k\right) - w\left(\hat{F}_{k+1}\right)\right)\right.\right. \\
&\left.\left. - \sum_{i=1}^{K} u_k\left(w\left(F_k\right) - w\left(F_{k+1}\right)\right)\right| \leq \epsilon\right) > 1 - \delta. \qquad (41)
\end{aligned}$$

Recall that w is locally Lipschitz continuous with constants $L_1, ...L_K$ at the points $F_1, ...F_K$. From a parallel argument to that in the proof of Proposition 8, it is easy to infer that

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_{i=1}^{K} u_k w(\hat{F}_{k+1}) - \sum_{i=1}^{K} u_k w(F_{k+1})\right| > \epsilon\right) \\
&\leq K\left(e^{-2n\rho^2} + e^{-2n\epsilon^2/(KLM)^2}\right).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_{i=1}^{K} u_k\left(w\left(\hat{F}_k\right) - w\left(\hat{F}_{k+1}\right)\right)\right.\right. \\
&\left.\left. - \sum_{i=1}^{K} u_k\left(w\left(F_k\right) - w\left(F_{k+1}\right)\right)\right| > \epsilon\right) \\
&\leq \mathbb{P}\left(\left|\sum_{i=1}^{K} u_k\left(w\left(\hat{F}_k\right)\right) - \sum_{i=1}^{K} u_k\left(w\left(F_k\right)\right)\right| > \epsilon/2\right) \\
&+ \mathbb{P}\left(\left|\sum_{i=1}^{K} u_k\left(w\left(\hat{F}_{k+1}\right)\right) - \sum_{i=1}^{K} u_k\left(w\left(F_{k+1}\right)\right)\right| > \epsilon/2\right) \\
&\leq 2K(e^{-2n\rho^2} + e^{-2n\epsilon^2/(KLM)^2}).
\end{aligned}$$

The claim in (41) now follows. □

B. Proofs for CPT-value optimization

1) Proofs for CPT-SPSA:

Lemma 4. *Let $\mathcal{F}_n = \sigma(\theta_m, m \leq n)$, $n \geq 1$. Then, for any $i = 1, \ldots, d$, we have almost surely,*

$$\left|\mathbb{E}\left[\frac{\overline{\mathbb{C}}_n^{\theta_n + \delta_n\Delta_n} - \overline{\mathbb{C}}_n^{\theta_n - \delta_n\Delta_n}}{2\delta_n\Delta_n^i}\middle|\mathcal{F}_n\right] - \nabla_i\mathbb{C}(X^{\theta_n})\right| \xrightarrow{n\to\infty} 0.$$

Proof. Notice that

$$\mathbb{E}\left[\frac{\overline{\mathbb{C}}_n^{\theta_n+\delta_n\Delta_n}-\overline{\mathbb{C}}_n^{\theta_n-\delta_n\Delta_n}}{2\delta_n\Delta_n^i}\mid\mathcal{F}_n\right] \tag{42}$$

$$=\mathbb{E}\left[\frac{\mathbb{C}(X^{\theta_n+\delta_n\Delta_n})-\mathbb{C}(X^{\theta_n-\delta_n\Delta_n})}{2\delta_n\Delta_n^i}\mid\mathcal{F}_n\right]+\mathbb{E}\left[\kappa_n\mid\mathcal{F}_n\right], \tag{43}$$

where $\kappa_n = \left(\frac{\psi^{\theta_n+\delta_n\Delta}-\psi^{\theta_n-\delta_n\Delta}}{2\delta_n\Delta_n^i}\right)$ is the estimation error arising out of the empirical distribution based CPT-value estimation scheme. From Corollary 1 and the fact that $\frac{1}{m_n^{\alpha/2}\delta_n}\to 0$ by assumption (A3), we have that

$$\mathbb{E}\kappa_n\to 0 \text{ a.s. as } n\to\infty.$$

Thus,

$$\mathbb{E}\left[\frac{\overline{\mathbb{C}}_n^{\theta_n+\delta_n\Delta_n}-\overline{\mathbb{C}}_n^{\theta_n-\delta_n\Delta_n}}{2\delta_n\Delta_n^i}\mid\mathcal{F}_n\right]$$

$$\xrightarrow{n\to\infty}\mathbb{E}\left[\frac{\mathbb{C}(X^{\theta_n+\delta_n\Delta_n})-\mathbb{C}(X^{\theta_n-\delta_n\Delta_n})}{2\delta_n\Delta_n^i}\mid\mathcal{F}_n\right]. \tag{44}$$

As in the case of regular SPSA, we simplify the RHS of (44) using suitable Taylor's expansions as follows:

$$\mathbb{E}\left[\frac{\mathbb{C}(X^{\theta_n+\delta_n\Delta_n})-\mathbb{C}(X^{\theta_n-\delta_n\Delta_n})}{2\delta_n\Delta_n^i}\mid\mathcal{F}_n\right]$$

$$=\nabla_i\mathbb{C}(X^{\theta_n})+\mathbb{E}\left[\sum_{j=1,j\neq i}^{N}\frac{\Delta_n^j}{\Delta_n^i}\right]\nabla_j\mathbb{C}(X^{\theta_n})+O(\delta_n^2)$$

$$=\nabla_i\mathbb{C}(X^{\theta_n})+O(\delta_n^2). \tag{45}$$

The first equality above follows from the fact that Δ_n is distributed according to a d-dimensional vector of symmetric, ±1-valued Bernoulli r.v.s and is independent of $\mathcal{F}_n$. The second inequality follows by observing that Δ_n^i is independent of Δ_n^j, for any $i,j=1,\ldots,d$, $j\neq i$.

The claim follows by using the fact that $\delta_n\to 0$ as $n\to\infty$. □

Proof. (***Theorem 1***)
We first rewrite the update rule (12) as follows: For $i=1,\ldots,d$,

$$\theta_{n+1}^i=\Pi_i\left(\theta_n^i+\gamma_n(\nabla_i\mathbb{C}(X^{\theta_n})+\beta_n+\xi_n)\right), \tag{46}$$

where

$$\beta_n=\mathbb{E}\left(\frac{(\overline{\mathbb{C}}_n^{\theta_n+\delta_n\Delta_n}-\overline{\mathbb{C}}_n^{\theta_n-\delta_n\Delta_n})}{2\delta_n\Delta_n^i}\mid\mathcal{F}_n\right)-\nabla_i\mathbb{C}(X^{\theta_n}),$$

$$\xi_n=\left(\frac{\overline{\mathbb{C}}_n^{\theta_n+\delta_n\Delta_n}-\overline{\mathbb{C}}_n^{\theta_n-\delta_n\Delta_n}}{2\delta_n\Delta_n^i}\right)$$

$$-\mathbb{E}\left(\frac{(\overline{\mathbb{C}}_n^{\theta_n+\delta_n\Delta_n}-\overline{\mathbb{C}}_n^{\theta_n-\delta_n\Delta_n})}{2\delta_n\Delta_n^i}\mid\mathcal{F}_n\right).$$

In the above, β_n is the bias in the gradient estimate due to SPSA and $\{\xi_n\}$ is a martingale difference sequence.

To prove the main claim, we list and verify assumptions (B1)-(B5), which are necessary to invoke Theorem 5.3.1 on pp. 191-196 of [30].

(B1): $\nabla\mathbb{C}(\cdot)$ is a continuous $\mathbb{R}^d$-valued function: holds by assumption in our setting.
(B2): The sequence $\{\beta_n, n\geq 0\}$ is a bounded random sequence with $\beta_n\to 0$ almost surely as $n\to\infty$: follows from Lemma 4.
(B3): The step-sizes $\gamma_n, n\geq 0$ satisfy $\gamma_n\to 0$ as $n\to\infty$ and $\sum_n\gamma_n=\infty$: holds by assumption (A3).
(B4): $\{\xi_n, n\geq 0\}$ is a sequence such that for any $\epsilon>0$,

$$\lim_{n\to\infty}P\left(\sup_{m\geq n}\left\|\sum_{k=n}^{m}\gamma_k\xi_k\right\|\geq\epsilon\right)=0. \tag{47}$$

We verify this assumption using arguments similar to those used in [11] for SPSA. Notice that

$$\mathbb{E}\left\|\xi_n\right\|^2\leq\mathbb{E}\left(\frac{\overline{\mathbb{C}}_n^{\theta_n+\delta_n\Delta_n}-\overline{\mathbb{C}}_n^{\theta_n-\delta_n\Delta_n}}{2\delta_n\Delta_n^i}\right)^2 \tag{48}$$

$$\leq\left(\left[\mathbb{E}\left(\frac{\overline{\mathbb{C}}_n^{\theta_n+\delta_n\Delta_n}}{2\delta_n\Delta_n^i}\right)^2\right]^{\frac{1}{2}}+\left[\mathbb{E}\left(\frac{\overline{\mathbb{C}}_n^{\theta_n-\delta_n\Delta_n}}{2\delta_n\Delta_n^i}\right)^2\right]^{\frac{1}{2}}\right)^2 \tag{49}$$

$$\leq\frac{\left[\mathbb{E}\left[\overline{\mathbb{C}}_n^{\theta_n+\delta_n\Delta_n}\right]^{2+2\alpha_2}\right]^{\frac{1}{1+\alpha_2}}+\left[\mathbb{E}\left[\overline{\mathbb{C}}_n^{\theta_n-\delta_n\Delta_n}\right]^{2+2\alpha_2}\right]^{\frac{1}{1+\alpha_2}}}{4\delta_n^2} \tag{50}$$

$$\leq\frac{C}{\delta_n^2},\text{ for some } C<\infty. \tag{51}$$

The inequality in (48) uses the fact that, for any random variable X, $\mathbb{E}\left\|X-E[X\mid\mathcal{F}_n]\right\|^2\leq\mathbb{E}X^2$. The inequality in (49) follows by the fact that $\mathbb{E}(X+Y)^2\leq\left((\mathbb{E}X^2)^{1/2}+(\mathbb{E}Y^2)^{1/2}\right)^2$. The inequality in (50) uses Hölder's inequality, with $\alpha_1,\alpha_2>0$ satisfying $\frac{1}{1+\alpha_1}+\frac{1}{1+\alpha_2}=1$ and the fact that $\mathbb{E}\left(\frac{1}{(\Delta_n^i)^{2+2\alpha_1}}\right)=1$ as Δ_n^i is a symmetric, ±1-valued Bernoulli r.v. The inequality in (51) follows by using the fact that $\mathbb{C}(X^\theta)$ is bounded a.s. for any parameter θ and the estimation error is bounded by Corollary 1. Thus, $\mathbb{E}\left\|\xi_n\right\|^2\leq\frac{C}{\delta_n^2}$ for some $C<\infty$.

Applying Doob's martingale inequality to the martingale difference $W_l:=\sum_{n=0}^{l-1}\gamma_n\xi_n$, $l\geq 1$, we obtain

$$\mathbb{P}\left(\sup_{l\geq k}\left\|\sum_{n=k}^{l}\gamma_n\xi_n\right\|\geq\epsilon\right)\leq\frac{1}{\epsilon^2}\sum_{n=k}^{\infty}\gamma_n^2\mathbb{E}\left\|\xi_n\right\|^2\leq\frac{dC}{\epsilon^2}\sum_{n=k}^{\infty}\frac{\gamma_n^2}{\delta_n^2},$$

and (47) follows by taking limits above and using (A3).
(B5): There exists a compact subset $\mathcal{K}$ which is the set of asymptotically stable equilibrium points for the ODE (13): To verify this assumption, observe that $\mathbb{C}(X^\theta)$ serves as a strict Lyapunov function for the ODE (13), since

$$\frac{d\mathbb{C}(X^\theta)}{dt}=\nabla\mathbb{C}(X^\theta)\dot{\theta}=\nabla\mathbb{C}(X^\theta)\check{\Pi}\left(-\nabla\mathbb{C}(X^\theta\right)\leq 0,$$

with strict inequality outside the set $\mathcal{K}'=\{\theta\mid\check{\Pi}_i\left(-\nabla\mathbb{C}(X^\theta)\right)=0,\forall i=1,\ldots,d\}$. Hence, the set $\mathcal{K}'$ serves as the asymptotically stable attractor for the ODE (13).

The claim follows from the Kushner-Clark lemma. □

2) Proofs for CPT-MPS:

Lemma 5. *The sequence of random variables* $\{\theta_n^*, n = 0, 1, \ldots\}$ *in Algorithm 2 converges w.p.1 as* $n \to \infty$.

Proof. Let $\mathcal{A}_n$ be the event that the first if statement is true within the *thresholding* step of Algorithm 2. Let $\mathcal{B}_n := \{\mathbb{C}(X^{\theta_n^*}) - \mathbb{C}(X^{\theta_{n-1}^*}) \leq \frac{\varepsilon}{2}\}$. Whenever $\mathcal{A}_n$ holds, we have $\overline{\mathbb{C}}_n^{\theta_n^*} - \overline{\mathbb{C}}_n^{\theta_{n-1}^*} \geq \varepsilon$ and hence, we obtain

$$
\begin{aligned}
&\mathbb{P}(\mathcal{A}_n \cap \mathcal{B}_n) \\
&\leq \mathbb{P}\left(\{\overline{\mathbb{C}}_n^{\theta_n^*} - \overline{\mathbb{C}}_{n-1}^{\theta_{n-1}^*} \geq \varepsilon\} \cap \left\{\mathbb{C}(X^{\theta_n^*}) - \mathbb{C}(X^{\theta_{n-1}^*}) \leq \frac{\varepsilon}{2}\right\}\right) \\
&\leq |\Lambda_n||\Lambda_{n-1}| \sup_{\theta,\theta' \in \Theta} \mathbb{P}\left(\{\overline{\mathbb{C}}_n^{\theta} - \overline{\mathbb{C}}_{n-1}^{\theta'} \geq \varepsilon\}\right. \\
&\qquad \left. \cap \left\{\mathbb{C}(X^{\theta}) - \mathbb{C}(X^{\theta'}) \leq \frac{\varepsilon}{2}\right\}\right) \\
&\leq |\Lambda_n||\Lambda_{n-1}| \sup_{\theta,\theta' \in \Theta} \left(\mathbb{P}\left(\overline{\mathbb{C}}_n^{\theta} - \mathbb{C}(X^{\theta}) \geq \frac{\varepsilon}{4}\right)\right. \\
&\qquad \left. + \mathbb{P}\left(\overline{\mathbb{C}}_{n-1}^{\theta'} - \mathbb{C}(X^{\theta'}) \geq \frac{\varepsilon}{4}\right)\right) \\
&\leq 4|\Lambda_n||\Lambda_{k-1}| e^{-\frac{m_n \epsilon^2}{8L^2M^2}}.
\end{aligned}
$$

From the foregoing, we have $\sum_{n=1}^{\infty} \mathbb{P}(\mathcal{A}_n \cap \mathcal{B}_n) < \infty$ since $m_n \to \infty$ as $n \to \infty$. Applying the Borel-Cantelli lemma, we obtain $\mathbb{P}(\mathcal{A}_n \cap \mathcal{B}_n \text{ i.o.}) = 0$. Hence, if $\mathcal{A}_n$ happens infinitely often, then $\mathcal{B}_n^c$ will also happen infinitely often and we have

$$
\begin{aligned}
&\sum_{n=1}^{\infty} \left[\mathbb{C}(X^{\theta_n^*}) - \mathbb{C}(X^{\theta_{n-1}^*})\right] = \sum_{\substack{n:\ \mathcal{A}_n \text{occurs}}} \left[\mathbb{C}(X^{\theta_n^*}) - \mathbb{C}(X^{\theta_{n-1}^*})\right] \\
&\quad + \sum_{\substack{n:\ \mathcal{A}_n^c \text{occurs}}} \left[\mathbb{C}(X^{\theta_n^*}) - \mathbb{C}(X^{\theta_{n-1}^*})\right] \\
&= \sum_{\substack{n: \\ \mathcal{A}_n \cap \mathcal{B}_n \\ \text{occurs}}} \left[\mathbb{C}(X^{\theta_n^*}) - \mathbb{C}(X^{\theta_{n-1}^*})\right] + \sum_{\substack{n: \\ \mathcal{A}_n \cap \mathcal{B}_n^c \\ \text{occurs}}} \left[\mathbb{C}(X^{\theta_n^*}) - \mathbb{C}(X^{\theta_{n-1}^*})\right] \\
&= \infty \quad \text{w.p.1, since } \varepsilon > 0.
\end{aligned}
$$

In the above, the first equality follows from the fact that if the else clause in thresholding step in Algorithm 2 is hit, then $\theta_n^* = \theta_{n-1}^*$. From the last equality above, we conclude that it is a contradiction because, $\mathbb{C}(X^{\theta}) < \mathbb{C}(X^{\theta^*})$ for any θ (since θ^* is the global maximum). The main claim now follows, since $\mathcal{A}_n$ can happen only a finite number of times. □

Proof. (***Theorem 2***)
Once we have established Lemma 5, the rest of the proof follows in an identical fashion as the proof of Corollary 4.18 of [12]. □

VI. Numerical Experiments

In this section as well as the next section, we show that the optimal CPT-value reacts differently to the change of parameters of the underlying distribution as compared to the optimal expected value. In other words, there are families of random variables $\{X_\theta, \theta \in \Theta\}$ where $\arg\max_\theta \mathbb{E}(X_\theta)$ is radically different from $\arg\max_\theta \mathbb{C}(X_\theta)$. This finding would make a case for specialized algorithms that optimize CPT-based criteria, since expected value optimizing algorithms cannot be used as surrogates.

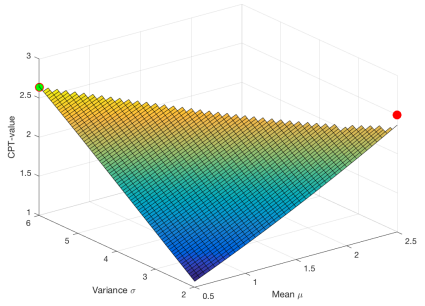


Fig. 3. CPT-value of normal distributed r.v.s with mean μ and variance σ parameters[1].

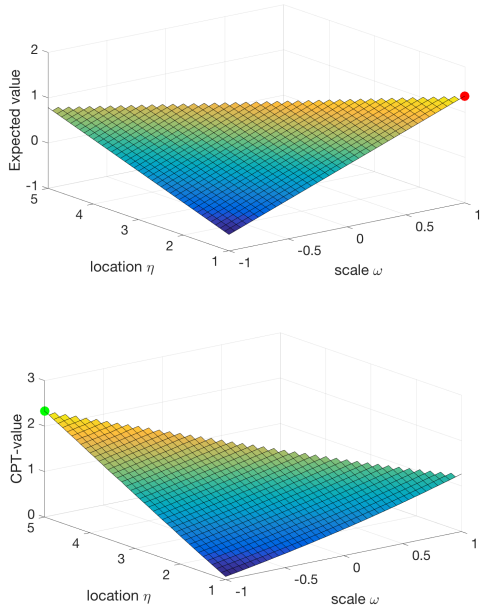


Fig. 4. Expected and CPT values of skewed normal distributed r.v.s with fixed shape $\alpha = 0.5$ and varying location ξ and scale ω parameters[1].

The CPT-value in this section is aligned with the form proposed in (1) and uses the following choices for utility and weight functions:

$$
\begin{aligned}
&u^+(x) = |x|^\sigma, \quad u^-(x) = \lambda|x|^\sigma, \\
&w^+(p) = \frac{p^{\eta_1}}{(p^{\eta_1} + (1-p)^{\eta_1})^{\frac{1}{\eta_1}}}, w^-(p) = \frac{p^{\eta_2}}{(p^{\eta_2} + (1-p)^{\eta_2})^{\frac{1}{\eta_2}}},
\end{aligned}
$$

where $\lambda = 0.25$, $\sigma = 0.88$, $\eta_1 = 0.61$ and $\eta_2 = 0.69$. The choices for σ and $w^+(\cdot)$, $w^-(\cdot)$ are based on the recommendations given by [8].

Since it is usually hard to obtain an analytical expression for the CPT-value, we use numerical integration via the trapezoidal rule. We consider two settings where the feasible region is triangle shaped over two distribution parameters. In each setting, the expected value optima is calculated analytically, while for the CPT-value, we perform a grid search, where the distance between points in the grid is 0.05.

Example 1. *We consider normal distributed r.v.s with mean* μ *and variance* σ. *As shown in Figure 3, the feasible region for* (μ, σ) *is the triangle with vertices* $(0.5, 2), (0.5, 6)$ *and*

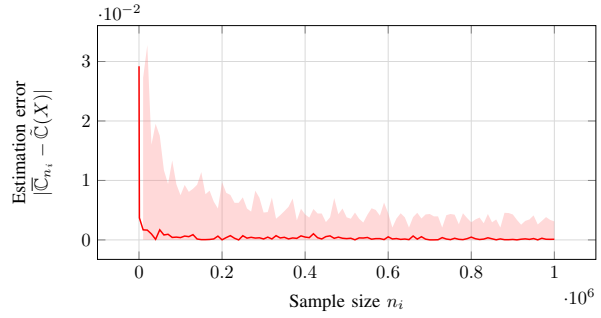


Fig. 5. Difference between CPT-estimate $\overline{\mathbb{C}}_{n_i}$ using Algorithm 1 and numerically integrated approximation $\tilde{\mathbb{C}}(X)$ to CPT-value $\mathbb{C}(X)$ of a skew normal distributed r.v. X with shape 2, location 2 and scale 1. The shaded bands denote the standard error calculated from ten independent simulations.

$(2.5, 2)$. *The expected value takes its maximum analytically at* $(2.5, 2)$*, while a numerical optimization of the CPT-value returned a maximum at* $(0.5, 6)$*, with corresponding CPT-value* 2.65*. The CPT value of the r.v.* $N(2.5, 2)$ *was* 2.37.

Example 2. *We consider skew normal distributed r.v.s* $sn(\xi, \omega, \alpha)$ *with location* ξ*, scale* ω *and shape* α*. The mean of* $X_{(\xi,\omega,\alpha)} \sim sn(\xi, \omega, \alpha)$ *is* $\xi + \omega\delta\sqrt{\frac{2}{\pi}}$*, while the variance is* $\omega^2(1 - \frac{2\delta^2}{\pi})$*, with* $\delta = \frac{\alpha}{1+\alpha^2}$*. With* $\alpha = 0.5$*, we set up the feasible region for* (ξ, ω) *to be the triangle with vertices* $(-1, 1), (1, 1)$ *and* $(-1, 5)$ *as shown in Figure 4. It turns out that the point* $(-1, 5)$ *returns the largest CPT-value, with* $\mathbb{C}(X_{-1,5,0.5,0.5}) = 2.30$*, while* $\mathbb{E}(X_{-1,5,0.5}) = 0.78$*. On the other hand, the point* $(1, 1)$ *has the largest expected value with* $\mathbb{E}(X_{1,1,0.5}) = 1.36$*, but the CPT value of the same r.v. is* 1.25*.*

We illustrate the rapid convergence of the estimator in Algorithm 1 for a skew normal distributed r.v. with location, scale and shape parameters set to $2, 1$ and 2, respectively. We conducted the experiment in 100 simulation phases indexed from 1 to 100. In each phase i, we generate i.i.d. estimators $\overline{\mathbb{C}}^j_{n_i}(X)$ with n_i samples of skew normal distributed r.v. X, where $j = 1, \ldots, 10$ corresponds to an independent simulation. The number of samples n_i in each phase i ranges from 100 to 10^6. For each phase i, we calculate the estimation error, which is the absolute difference between $\overline{\mathbb{C}}_{n_i} = \frac{1}{10}\sum_{j=1}^{10} \mathbb{C}^j_{n_i}$ and the numerically integrated CPT-value. Figure 5 presents the estimation error with standard error bars for each n_i.

VII. Traffic Control Application

We consider a traffic signal control application where the aim is to improve the road user experience by an adaptive traffic light control (TLC) algorithm. We optimize the CPT-value of the delay experienced by road users, since CPT realistically captures the attitude of the road users towards delays. It is assumed that the CPT functional's parameters (u, w) are given (usually, these are obtained by observing human behavior). The experiments are performed using the GLD traffic simulator [31], and the implementation is available at https://bitbucket.org/prashla/rl-gld.

[1]The red dot is the expected value optima that is calculated analytically, while the green dot is the CPT-value optima.

We consider a road network with $\mathcal{N}$ signalled lanes that are spread across junctions and $\mathcal{M}$ paths, where each path connects (uniquely) two edge nodes, from which the traffic is generated (see Figure 6). At any instant n, let q^i_n and t^i_n denote the queue length and elapsed time since the lane turned red, for any lane $i = 1, \ldots, \mathcal{N}$. Let $d^{i,j}_n$ denote the delay experienced by jth road user on ith path, for any $i = 1, \ldots, \mathcal{M}$ and $j = 1, \ldots, n_i$, where n_i denotes the number of road users on path i. We specify the various components of the traffic control MDP below. The state $s_n = (q^1_n, \ldots, q^{\mathcal{N}}_n, t^1_n, \ldots, t^{\mathcal{N}}_n, d^{1,1}_n, \ldots, d^{\mathcal{M},n_{\mathcal{M}}}_n)^\intercal$ is a vector of lane-wise queue lengths, elapsed times and pathwise delays. Any combination of traffic lights that can simultaneously be switched to green constitutes an action in the MDP.

We consider Boltzmann policies that have the form

$$\pi_\theta(s, a) = \frac{e^{\theta^\top \phi_{s,a}}}{\sum_{a' \in \mathcal{A}(s)} e^{\theta^\top \phi_{s,a'}}}, \quad \forall s \in \mathcal{S},\ \forall a \in \mathcal{A}(s),$$

with features $\phi_{s,a}$ as described in Section V-B of [32]. For any policy θ, let X^θ_i be the delay r.v. and μ^θ_i the proportion of road users along path i, for $i = 1, \ldots, \mathcal{M}$. Any road user along path i will evaluate the delay (s)he experiences in a manner that is captured well by CPT. An important component of CPT is to employ a reference point to calculate gains and losses. In our setting, we use pathwise delays, say B_i for path i, obtained from a pre-timed TLC (cf. the Fixed TLCs in [33]) as the reference point. If the delay of any TLC algorithm is less than that of pre-timed TLC, then the (positive) difference in delays is perceived as a gain and in the complementary case, the delay difference is perceived as a loss. Thus, the CPT-value $\mathbb{C}(B_i - X_i)$ for any path i in (52) is to be understood as a *differential delay gain* w.r.t. B_i. Now, the objective is to maximize the weighted sum of CPT-values across paths, i.e.,

$$\max_{\theta \in \Theta} \text{CPT}(X^\theta_1, \ldots, X^\theta_{\mathcal{M}}) = \sum_{i=1}^{\mathcal{M}} \mu^\theta_i \mathbb{C}(B_i - X^\theta_i), \tag{52}$$

where Θ is the d-dimensional hypercube formed by intervals $[0.1, 1.0]$ in each dimension. The rationale behind the objective above is that CPT-value $\mathbb{C}(B_i - X^\theta_i)$ would capture the road user experience/satisfaction for each path i and the goal is to maximize the *average satisfaction* over all paths.

For the sake of comparison, we consider the traditional objective of minimizing the overall average delay, i.e.,

$$\min_{\theta \in \Theta} \text{AVG}(X^\theta_1, \ldots, X^\theta_{\mathcal{M}}) = \sum_{i=1}^{\mathcal{M}} \mu^\theta_i \mathbb{E}(X^\theta_i). \tag{53}$$

In comparison to CPT objective, the above does not incorporate baseline delays, makes no distinction between gains and losses via utility functions and does not distort probabilities.

We implement the following TLC algorithms:

CPT-SPSA: This is a first-order algorithm that solves (52) using SPSA-based gradient estimates and Algorithm 1 for estimating CPT-value $\mathbb{C}(B_i - X_i)$ for each path $i = 1, \ldots, \mathcal{M}$, with $d^{i,j}_n, j = 1, \ldots, n_i$ as the samples.

AVG-SPSA: This is SPSA-based first-order algorithm that solves (53), while using sample averages of the delays to estimate the expected delay $\mathbb{E}(X_i)$ for each path $i = 1, \ldots, \mathcal{M}$.

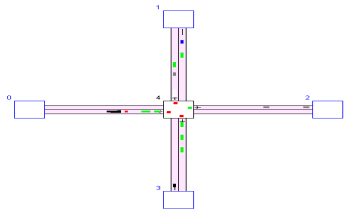
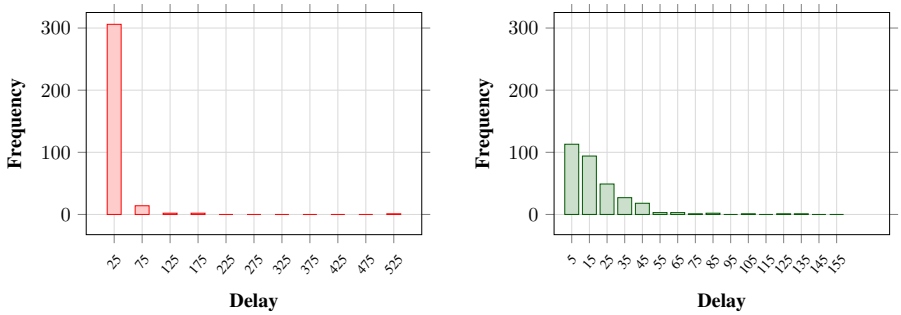
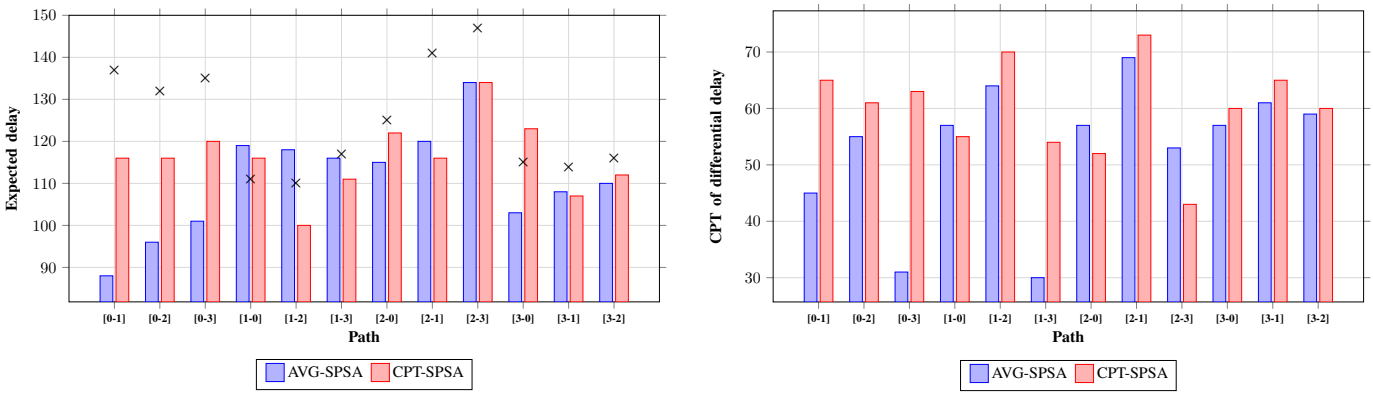


Fig. 6. Snapshot of the road network from GLD simulator. The figure shows four edge nodes that generate traffic, one traffic light and two-laned roads carrying automobiles.



(a) AVG-SPSA (b) CPT-SPSA

Fig. 7. Histogram of the sample delays for the path from node 0 to 1 (see Figure 6) for AVG-SPSA that minimizes overall expected delay and CPT-SPSA that maximizes CPT-value of differential delay.



(a) Expected delay (path-wise). The × marks indicate baseline delays.

(b) CPT of differential delay (pathwise).

Fig. 8. AVG and CPT values for two algorithms on each of the 12 paths in Figure 6: AVG-SPSA minimizes overall expected delay (see (53)), while CPT-SPSA maximizes CPT-value of differential delay (see (52)).

TABLE I
AVG AND CPT-VALUE ESTIMATES FOR AVG-SPSA AND CPT-SPSA.

	AVG-value	**CPT-value**
AVG-SPSA	**111.67**	53.31
CPT-SPSA	116.21	**59.91**

Recent works in [34], [35] consider the problem of traffic light control with parameterized policies based on underlying queue lengths and elapsed times. However, they do not consider a CPT-based objective and apply a perturbation analysis approach that imposes structural restrictions on the underlying objective.

The underlying CPT-value $\mathbb{C}(X_i), \forall i$ follows the exact form as in section VI, except here we set $\lambda = 2.25$. The choices for λ, σ, η_1 and η_2 are based on median estimates given by [8] and have been used earlier in a traffic application (see [36]). For all the algorithms, motivated by standard guidelines (see [37]), we set $\delta_n = 1.9/n^{0.101}$ and $a_n = 1/(n + 50)$. The initial point θ_0 is the d-dimensional vector of ones and $\forall i$, the operator Γ_i keeps the iterate θ_i within $[0.1, 1.0]$.

The experiments involve two phases: first, a training phase where we run each algorithm for 500 iterations, with each iteration involving two perturbed simulations. Each simulation involves running the traffic simulator with a fixed policy parameter for 5000 steps and this corresponds to approximately 4000 delay samples. The training phase is followed by a test phase where we fix the policy obtained at the end of training and then run the traffic simulator with the aforementioned parameter for 5000 steps. The results presented are averages over ten independent simulations.

Table I presents the overall AVG and CPT-values for AVG-SPSA and CPT-SPSA, while Figures 8(a)–8(b) present the expected delay and CPT of differential delay for each of the 12 paths in Figure 6. We observe that AVG-SPSA exhibits a lower AVG-value, while CPT-SPSA shows a higher CPT-value. Further, from Figure 7 that presents the histograms of the delays for the path from 0 to 1, we observe that CPT-SPSA results in a strategy that avoids high delays at the cost of a slightly higher average delay, whereas AVG-SPSA occasionally incurs delays significantly larger than the average delay.

VIII. CONCLUSIONS

CPT has been a very popular paradigm for quantifying human preferences among psychologists/economists, and this work is the first step in incorporating CPT-based criteria into a stochastic optimization framework. Estimation and optimization of the CPT-based value is challenging. For estimat-

ing the CPT-value, we proposed a quantile-based estimation scheme and for maximizing the CPT-value, we adapted the SPSA [11] and MPS [12] algorithms. We provided theoretical convergence guarantees for all the proposed algorithms and illustrated the usefulness of our algorithms for optimizing CPT-based criteria in a traffic signal control application.

REFERENCES

[1] J. Von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*. Princeton: Princeton University Press, 1944.

[2] P. Fishburn, *Utility Theory for Decision Making*. Wiley, New York, 1970.

[3] M. Allais, "Le comportement de l'homme rationel devant le risque: Critique des postulats et axioms de l'ecole americaine," *Econometrica*, vol. 21, pp. 503–546, 1953.

[4] D. Ellsberg, "Risk, ambiguity and the Savage's axioms," *The Quarterly Journal of Economics*, vol. 75, no. 4, pp. 643–669, 1961.

[5] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica: Journal of the Econometric Society*, pp. 263–291, 1979.

[6] C. Starmer, "Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk," *Journal of economic literature*, pp. 332–382, 2000.

[7] J. Quiggin, *Generalized Expected Utility Theory: The Rank-dependent Model*. Springer Science & Business Media, 2012.

[8] A. Tversky and D. Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, vol. 5, no. 4, pp. 297–323, 1992.

[9] L. A. Prashanth, C. Jie, M. C. Fu, S. I. Marcus, and C. Szepesvári, "Cumulative prospect theory meets reinforcement learning: Prediction and control," in *International Conference on Machine Learning*, 2016, pp. 1406–1415.

[10] L. A. Wasserman, *All of Nonparametric Statistics*. Springer, 2015.

[11] J. C. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Trans. Auto. Cont.*, vol. 37, no. 3, pp. 332–341, 1992.

[12] H. S. Chang, J. Hu, M. C. Fu, and S. I. Marcus, *Simulation-based Algorithms for Markov Decision Processes*. Springer, 2013.

[13] H. Markowitz, "Portfolio selection," *The journal of finance*, vol. 7, no. 1, pp. 77–91, 1952.

[14] K. J. Arrow, *Essays in the Theory of Risk Bearing*. Chicago, IL: Markham, 1971.

[15] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *Journal of risk*, vol. 2, pp. 21–42, 2000.

[16] M. Sobel, "The variance of discounted Markov decision processes," *Journal of Applied Probability*, pp. 794–802, 1982.

[17] J. Filar, L. Kallenberg, and H. Lee, "Variance-penalized Markov decision processes," *Mathematics of Operations Research*, vol. 14, no. 1, pp. 147–161, 1989.

[18] S. Mannor and J. N. Tsitsiklis, "Algorithmic aspects of mean–variance optimization in Markov decision processes," *European Journal of Operational Research*, vol. 231, no. 3, pp. 645–653, 2013.

[19] V. Borkar and R. Jain, "Risk-constrained Markov decision processes," in *IEEE Conference on Decision and Control*, 2010, pp. 2664–2669.

[20] L. A. Prashanth, "Policy gradients for CVaR-constrained MDPs," in *Algorithmic Learning Theory*, 2014, pp. 155–169.

[21] K. Lin, "Stochastic systems with cumulative prospect theory," Ph.D. dissertation, University of Maryland, College Park, 2013.

[22] A. Gopalan, L. Prashanth, M. Fu, and S. Marcus, "Weighted bandits or: How bandits learn distorted values that are not expected," in *AAAI Conference on Artificial Intelligence*, 2017, pp. 1941–1947.

[23] N. C. Barberis, "Thirty years of prospect theory in economics: A review and assessment," *Journal of Economic Perspectives*, pp. 173–196, 2013.

[24] D. Prelec, "The probability weighting function," *Econometrica*, pp. 497–527, 1998.

[25] S. Bhatnagar, H. L. Prasad, and L. Prashanth, *Stochastic Recursive Algorithms for Optimization*. Springer, 2013, vol. 434.

[26] V. Borkar, *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, 2008.

[27] J. Baxter and P. L. Bartlett, "Infinite-horizon policy-gradient estimation," *Journal of Artificial Intelligence Research*, vol. 15, pp. 319–350, 2001.

[28] B. Póczos, Y. Abbasi-Yadkori, C. Szepesvári, R. Greiner, and N. Sturtevant, "Learning when to stop thinking and do something!" in *International Conference on Machine Learning*, 2009, pp. 825–832.

[29] B. Yu, "Assouad, Fano, and Le Cam," in *Festschrift for Lucien Le Cam*. Springer, 1997, pp. 423–435.

[30] H. Kushner and D. Clark, *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer-Verlag, 1978.

[31] M. Wiering, J. Vreeken, J. van Veenen, and A. Koopman, "Simulation and optimization of traffic in a city," in *IEEE Intelligent Vehicles Symposium*, June 2004, pp. 453–458.

[32] L. A. Prashanth and S. Bhatnagar, "Threshold tuning using stochastic optimization for graded signal control," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 9, pp. 3865–3880, 2012.

[33] ——, "Reinforcement learning with function approximation for traffic signal control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 412–421, 2011.

[34] J. L. Fleck, C. G. Cassandras, and Y. Geng, "Adaptive quasi-dynamic traffic light control," *IEEE Transactions on Control Systems Technology*, vol. 24, no. 3, pp. 830–842, 2016.

[35] Y. Geng and C. G. Cassandras, "Multi-intersection traffic light control with blocking," *Discrete Event Dynamic Systems*, vol. 25, no. 1-2, pp. 7–30, 2015.

[36] S. Gao, E. Frejinger, and M. Ben-Akiva, "Adaptive route choices in risky traffic networks: A prospect theory approach," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 5, pp. 727–740, 2010.

[37] J. C. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. John Wiley & Sons, 2005, vol. 65.

Cheng Jie is a fourth year Ph.D. candidate in Department of Mathematics, University of Maryland - College Park. His research interests are in applied probability, stochastic optimization and machine learning.

Prashanth L.A. is currently an Assistant Professor in the Department of Computer Science and Engg., Indian Institute of Technology Madras. He received his Masters and Ph.D. degrees in Computer Science and Automation from Indian Institute of Science, in 2008 and 2013, respectively. He was awarded the third prize for his Ph.D. dissertation, by the IEEE Intelligent Transportation Systems Society (ITSS). His research interests are in reinforcement learning, stochastic optimization and multi-armed bandits, with applications in transportation, networks and recommendation systems.

Michael Fu received degrees in mathematics and EECS from MIT in 1985 and a Ph.D. in applied math from Harvard in 1989. Since 1989, he has been at the University of Maryland, College Park, currently holding the Smith Chair of Management Science. He also served as the Operations Research Program Director at the National Science Foundation. His research interests include simulation optimization and stochastic gradient estimation. He is a Fellow of the Institute for Operations Research and the Management Sciences.

Steve Marcus received the B.A. degree from Rice University in 1971 and the S.M. and Ph.D. degrees from M.I.T. in 1972 and 1975, respectively. From 1975 to 1991, he was with the Department of Electrical and Computer Engineering at the University of Texas at Austin. In 1991, he joined the University of Maryland, College Park, as Professor in the Electrical and Computer Engineering Department and the Institute for Systems Research. He was Director of the Institute for Systems Research from 1991 to 1996 and Chair of the Electrical and Computer Engineering Department from 2000 to 2005. Currently, his research is focused on stochastic control, estimation, hybrid systems, and optimization. He is a Fellow of the Society for Industrial and Applied Mathematics.

Csaba Szepesvari (PhD'99) is currently a Professor at the Department of Computing Science of the University of Alberta and a Principal Investigator of the Alberta Machine Intelligence Institute. He has published two books and nearly 200 papers in academic journals and conferences. He serves as an action editor of the Journal of Machine Learning Research and the Machine Learning Journal. He served as a co-chair for the 2014 Conference on Learning Theory, and for the Algorithmic Learning Theory Conference in 2011, as well as a senior PC member for NIPS, ICML, AISTATS, AAAI and IJCAI for many years. His interest is in designing principled learning algorithms for agents that learn while controlling their environment.