Appendix for A New Decision-dependent DRO Framework Considering Indirect Network Effects for Hydrogen Supply Infrastructure Expansion Planning

Haoran Deng, Bo Yang, Dafeng Zhu, Cailian Chen, Xinping Guan, and Dipti Srinivasan

Appendix

A. Investment Constraints

Similar to the investment constraints (3) for HRS, that of HS, P2G, PIP are as follows.

$$w_{i1}^{O,HS} = w_{i1}^{HS} \tag{47a}$$
$$w_{is}^{O,HS} = w_{i(s-1)}^{O,HS} + w_{is}^{HS} - w_{is}^{HS,retire} \tag{47b}$$
$$w_{is}^{HS,retire} \leqslant w_{i(s-1)}^{O,HS} \tag{47c}$$
$$w_{is}^{HS,retire} + w_{i(s-1)}^{HS,retire} \leqslant 1 \tag{47d}$$
$$w_{i(s-1)}^{HS} = w_{i(s+R^{HS}-1)}^{HS,retire} (1 \leqslant R^{HS} \leqslant S, s \geq 2) \tag{47e}$$

$$w_{i1}^{O,P2G} = w_{i1}^{P2G} \tag{48a}$$
$$w_{is}^{O,P2G} = w_{i(s-1)}^{O,P2G} + w_{is}^{P2G} - w_{is}^{P2G,retire} \tag{48b}$$
$$w_{is}^{P2G,retire} \leqslant w_{i(s-1)}^{O,P2G} \tag{48c}$$
$$w_{is}^{P2G,retire} + w_{i(s-1)}^{P2G,retire} \leqslant 1 \tag{48d}$$
$$w_{i(s-1)}^{P2G} = w_{i(s+R^{P2G}-1)}^{P2G,retire} (1 \leqslant R^{P2G} \leqslant S, s \geq 2) \tag{48e}$$

$$w_{l1}^{O} = w_{l1} \tag{49a}$$
$$w_{ls}^{O} = w_{l(s-1)}^{O} + w_{ls} - w_{ls}^{retire} \tag{49b}$$
$$w_{ls}^{retire} \leqslant w_{l(s-1)}^{O} \tag{49c}$$
$$w_{ls}^{retire} + w_{l(s-1)}^{retire} \leqslant 1 \tag{49d}$$
$$w_{l(s-1)} = w_{l(s+R^{L}-1)}^{retire} (1 \leqslant R^{L} \leqslant S, s \geq 2) \tag{49e}$$

Next, the constraint (4) which represent the relation between HS and HRS planning decisions is linearized. Constraints (4) is equivalent to $w_{is}^{O,HS} = -\max\{-1, -\sum_{j=1}^{J_i} w_{ijs}^{O,HRS}\}$. Let $y_{is} = \max\{-1, -\sum_{j=1}^{J_i} w_{ijs}^{O,HRS}\}$ and $\hat{y}_{is} = y_{is} + 1 = \max\{0, 1 - \sum_{j=1}^{J_i} w_{ijs}^{O,HRS}\}$. We introduce a binary variable θ_{is}, and obtain its equivalent form using the Big-M method:

$$(1 - \sum_{j=1}^{J_i} w_{ijs}^{O,HRS}) - (1 - \theta_{is})M \leqslant \hat{y}_{is} \tag{50a}$$
$$(1 - \sum_{j=1}^{J_i} w_{ijs}^{O,HRS}) + (1 - \theta_{is})M \geqslant \hat{y}_{is} \tag{50b}$$
$$(\theta_{is} - 1)M \leqslant 1 - \sum_{j=1}^{J_i} w_{ijs}^{O,HRS} \leqslant \theta_{is}M \tag{50c}$$
$$0 \leqslant \hat{y}_{is} \leqslant \theta_{is}M \tag{50d}$$

B. Hydrogen Demand and Transportation Network Constraints

In the case of (10a), the bounds of continuous variables $x_{its}^{HFCV,G}$ and g_{is} are $[0, x_i^{G,\max}]$ and $[g_{is}^{\min}, R^{HFCV}]$, respectively. We introduce a binary variables z_{isk}, and discretize the variable g_{is} using 2^K discrete units, which can be expressed by the binary expansion scheme [1] as

$$g_{is} = g_{is}^{\min} + \left(R^{HFCV} - g_{is}^{\min}\right) \Big/ 2^K \sum_{k=1}^{K} 2^{k-1} z_{isk} \tag{51}$$

where the constant K is a nonnegative integer. And then $x_{its}^{HFCV,G} g_{is}$ can be rewritten as

$$g_{its}^{HFCV} = g_{is}^{\min} x_{its}^{HFCV,G} + \frac{\left(R^{HFCV} - g_{is}^{\min}\right)}{2^K} \sum_{k=1}^{K} 2^{k-1} z_{isk} x_{its}^{HFCV,G} \tag{52}$$

which includes the form that multiplies a binary variable and a continuous variable. Let $v_{itsk} = z_{isk} x_{its}^{HFCV,G}$, and (10a) can be equivalently replaced by the functions as follows.

$$g_{its}^{HFCV} = g_{is}^{\min} x_{its}^{HFCV,G} + \frac{(R^{HFCV} - g_{is}^{\min})}{2^K} \sum_{k=1}^{K} 2^{k-1} v_{itsk} \tag{53a}$$
$$0 \leqslant x_{its}^{HFCV,G} - v_{itsk} \leqslant x_i^{G,\max}(1 - z_{isk}) \tag{53b}$$
$$0 \leqslant v_{itsk} \leqslant x_i^{G,\max} z_{isk} \tag{53c}$$

For the UE constraint (54) in the traffic assignment model, it represents a stable traffic flow state when travel times are equal and no greater than those of unselected paths for a particular O-D pair. In the user equilibrium pattern, the user equilibrium can capture the selfish behavior of vehicle users. Notably, we default that only the original traffic flow without redistribution satisfies the UE constraint, and the redistributed traffic flow is not required to satisfy it.

$$0 \leqslant f_{pts}^{od} \perp (c_{pts}^{od} - c_{pts}^{od,opt}) \geqslant 0 \tag{54}$$

Corresponding author: Bo.Yang

H. Deng, B. Yang, D. Zhu, C. Chen, and X. Guan are with the Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China (email: denghaoran@sjtu.edu.cn; bo.yang@sjtu.edu.cn; dafez@kth.se; cailianchen@sjtu.edu.cn; xpguan@sjtu.edu.cn).

D. Srinivasan is with the Department of Electrical & Computer Engineering, National University of Singapore, 117576, Singapore (email: dipti@nus.edu.sg).

where $f_{pts}^{od} = f_{pts}^{od,HFCV,G} + f_{pts}^{od,OTV}$. The lowest travel time on path p is $c_{pts}^{od} = \sum_{l\in\mathcal{L}_T} t_{l,t}^{G}(x_{lts}^{G})\delta_{l,p}^{od} = \sum_{l\in\mathcal{L}_T} t_l^0 \left[1+0.15(x_{lts}^{G}/C_l^{\max})^4\right]\delta_{l,p}^{od}$, which can be linearized by the piecewise linear method of hydrogen pipeline flow constraints in ***Appendix. C***. The UE principle modeling of transportation network has been widely used. Refer to [2] to linearize it as

$$0 \leqslant f_{pts}^{od} \leqslant M(1-\vartheta_{pts}^{od}) \tag{55a}$$
$$0 \leqslant c_{pts}^{od} - c_{pts}^{od,opt} \leqslant M\vartheta_{pts}^{od} \tag{55b}$$

For the linearization method of Eq. (15) and (16), we mark the function $G_s\left(g_{re,s}^{B},\rho_{re,s}\right) = g_{re,s}^{B} - \frac{\alpha}{\sqrt{2}}\left(\rho_{re,s}\right)^{-\frac{1}{4}} = 0$, and set the variable $\boldsymbol{o} = \left(g_{re,s}^{B},\rho_{re,s}\right)$. According to the Taylor series expansion of $G_s\left(g_{re,s}^{B},\rho_{re,s}\right)$, we have:

$$\begin{aligned}&\boldsymbol{G}(\boldsymbol{o}^{(1)}) = \boldsymbol{G}(\boldsymbol{o}^{(0)}+\varDelta\boldsymbol{o}^{(0)}) = \boldsymbol{G}(\boldsymbol{o}^{(0)}) + \\ &\boldsymbol{G}'(\boldsymbol{o}^{(0)})\varDelta\boldsymbol{o}^{(0)} + \boldsymbol{G}''(\boldsymbol{o}^{(0)})\frac{(\varDelta\boldsymbol{o}^{(0)})^2}{2!} + \cdots = 0\end{aligned} \tag{56}$$

Eq. (56) can be approximated by first-order Taylor series as

$$\boldsymbol{G}(\boldsymbol{o}^{(1)}) = \boldsymbol{G}(\boldsymbol{o}^{(0)}) + \boldsymbol{G}'(\boldsymbol{o}^{(0)})\varDelta\boldsymbol{o}^{(0)} = 0 \tag{57}$$

Then there is the following iteration relation

$$\begin{cases}\boldsymbol{J}^{(n)}\varDelta\boldsymbol{o}^{(n)} = -\boldsymbol{G}\left(\boldsymbol{o}^{(n)}\right)\\ \boldsymbol{o}^{(n+1)} = \boldsymbol{o}^{(n)} + \varDelta\boldsymbol{o}^{(n)}\end{cases} \tag{58}$$

where $\boldsymbol{J}^{(n)}$ is the Jacobian matrix in the n-th iteration, and it is given as $\boldsymbol{J}^{(n)} = \frac{\partial G}{\partial \boldsymbol{o}} \mid_{\boldsymbol{o}^{(n)}}$. We use $\boldsymbol{o}^{(0)}$ as the initial value to further estimate the value of $\boldsymbol{o}$ in the next iteration.

Next, the coefficient Jacobian matrix are as follows:

$$\begin{cases}\frac{\partial G_s}{\partial g_{re,s}^{B}} \mid_{\boldsymbol{o}^{(n)}} = \boldsymbol{I}\\ \frac{\partial G_s}{\partial \rho_{re,s}} \mid_{\boldsymbol{o}^{(n)}} = \frac{\alpha}{4\sqrt{2}}\left(\rho_{re,s}\right)^{-\frac{5}{4}} \mid_{\boldsymbol{o}^{(n)}} = \boldsymbol{J}_{\rho}^{(n)}\end{cases} \tag{59}$$

The iterative relation (58) is specified as:

$$\begin{cases}[\boldsymbol{I},\boldsymbol{J}_{\rho}^{(n)}]\begin{bmatrix}\varDelta g_{re,s}^{B(n)}\\ \varDelta\rho_{re,s}^{(n)}\end{bmatrix} = -\boldsymbol{G}(g_{re,s}^{B(n)},\rho_{re,s}^{(n)})\\ \begin{bmatrix}g_{re,s}^{B(n+1)}\\ \rho_{re,s}^{(n+1)}\end{bmatrix} = \begin{bmatrix}g_{re,s}^{B(n)}\\ \rho_{re,s}^{(n)}\end{bmatrix} + \begin{bmatrix}\varDelta g_{re,s}^{B(n)}\\ \varDelta\rho_{re,s}^{(n)}\end{bmatrix}\end{cases} \tag{60}$$

The termination condition of this iteration process is $\|\varDelta\boldsymbol{o}\| \leqslant \varepsilon_G$, where ε_G is a small number.

C. Hydrogen Network and Power System Constraints

The specific hydrogen network constraints are as follows. Among them, constraints (61a)-(61e) are represented as the hydrogen pipeline flow model and linepack model. (61f)-(61i) denotes the HS constraints. The hydrogen balance constraints for each node are shown in (61j)-(61m).

$$\begin{aligned}&-(1-\min\{1,w_{ls}^{O}\})M \leqslant g_{lts}^2 - \phi_l(pr_{s(l),ts}^2 - pr_{e(l),ts}^2)\cdot\\ &\operatorname{sgn}(pr_{s(l),ts},pr_{e(l),ts}) \leqslant (1-\min\{1,w_{ls}^{O}\})M\end{aligned} \tag{61a}$$
$$\underline{pr} \leqslant pr_{its} \leqslant \overline{pr}(l\in\mathcal{L}), g_{lts} = 0.5(g_{lts}^{pipout} + g_{lts}^{pipin}) \tag{61b}$$
$$0 \leqslant g_{lts}, g_{lts}^{pipout}, g_{lts}^{pipin} \leqslant w_{ls}^{O}\overline{F_l} \tag{61c}$$
$$e_{lts}^{pip} = e_{l(t-1)s}^{pip} + \varDelta t(g_{lts}^{\text{pipin}} - g_{lts}^{\text{pipout}}) \tag{61d}$$
$$0 \le e_{lts}^{pip} \le \overline{E_l}w_{ls}^{O}, e_{lTs}^{pip} = e_{l0s}^{pip} \tag{61e}$$
$$H_{its} = H_{i(t-1)s} + \eta_{HC}g_{its}^{ch}\Delta t - g_{its}^{dis}\Delta t/\eta_{HD} \tag{61f}$$
$$\underline{H}_i^{HS}\min\{1,w_{is}^{O,HS}\} \le g_{its}^{ch}, g_{its}^{dis} \le \overline{H}_i^{HS}\min\{1,w_{is}^{O,HS}\} \tag{61g}$$
$$0 \le H_{its} \le \sum\nolimits_{s=1}^{s} h_{is}^{HS}(w_{is}^{HS} - w_{is}^{HS,retire}) \tag{61h}$$
$$H_{i0s} = H_{iTs} \tag{61i}$$
$$g_{its}^{M} + g_{its}^{P2G} + g_{its}^{dis} - g_{its}^{ch} = g_{its}^{D} \tag{61j}$$
$$g_{its}^{M} = \sum\nolimits_{l\in\mathcal{L}_e(i)} g_{lts}^{pipout} - \sum\nolimits_{l\in\mathcal{L}_s(i)} g_{lts}^{pipin} \tag{61k}$$
$$0 \le g_{its}^{M} \le \overline{H}_i^{M}\min\{1,w_{is}^{O,HS}\} \tag{61l}$$
$$g_{its}^{P2G} = \eta_{P2G}\chi_{P2G}p_{its}^{P2G} \tag{61m}$$

The piecewise linear method [3] is applied to obtain the approximate linearization of g_{lts}^2 as follows.

$$g_{lts} = \sum\nolimits_{k=1}^{K}\varDelta g_{lts}^{(k)} \quad (l\in\mathcal{L}) \tag{62a}$$
$$0 \leqslant \varDelta g_{lts}^{(k)} \leqslant \overline{F_l}/K \tag{62b}$$
$$h(g_{lts}) = g_{lts}^2 = \sum\nolimits_{k=1}^{K}\lambda_{G,l}^{(k)}\varDelta g_{lts}^{(k)} \tag{62c}$$
$$\lambda_{G,l}^{(k)} = \frac{h\left(k\left(\frac{\bar{F}_l}{K}\right)\right) - h\left((k-1)\left(\frac{\bar{F}_l}{K}\right)\right)}{\bar{F}_l/K} = \frac{2k-1}{K}\bar{F}_l \tag{62d}$$

D. Wasserstein-based Ambiguity Set

The Wasserstein distance is defined as the minimum transportation cost of moving from probability $\mathbb{P}_1$ to $\mathbb{P}_2$. The specific arguments can be found in [4]. Mathematically,

$$d_s^{od}(\mathbb{P}_1,\mathbb{P}_2) = \inf_{\Pi\in\mathcal{Q}(\Xi,\Xi)}\left\{\begin{array}{l}\int_{\Xi^2}\left\|\boldsymbol{q}_{1,s}^{od} - \boldsymbol{q}_{2,s}^{od}\right\|\Pi\left(d\boldsymbol{q}_{1,s}^{od}\times d\boldsymbol{q}_{2,s}^{od}\right):\\ \Pi\left(d\boldsymbol{q}_{1,s}^{od},\Xi\right) = \mathbb{P}_1\left(d\boldsymbol{q}_{1,s}^{od}\right),\\ \Pi\left(\Xi,d\boldsymbol{q}_{2,s}^{od}\right) = \mathbb{P}_2\left(d\boldsymbol{q}_{2,s}^{od}\right)\end{array}\right\} \tag{63}$$

where $\mathcal{Q}(\mathbb{P}_1,\mathbb{P}_2)$ is the set of all joint distributions of $\left(\boldsymbol{q}_{1,s}^{od},\boldsymbol{q}_{2,s}^{od}\right)$ supported on Ξ^2 with marginals $(\mathbb{P}_1,\mathbb{P}_2)$. We construct the following Wasserstein ambiguity set to quantify differences between the empirical data-based distribution and the real unknown data-generating distribution:

$$\mathcal{P}(\hat{\mathbb{P}}_{N^Q},\boldsymbol{\varepsilon}) = \{\mathbb{P}\in\mathcal{P}(\varXi)\,\Big|\, d_s^{od}(\mathbb{P},\hat{\mathbb{P}}_{N^Q}) \leqslant \varepsilon_s^{od}\} \tag{64}$$

where $\hat{\mathbb{P}}_{N^Q}(\boldsymbol{m}) = \frac{1}{N^Q}\sum_{n\in N^Q}\mathbb{1}(\hat{\boldsymbol{q}}^n = \boldsymbol{m})\,(\forall\boldsymbol{m}\in\Xi)$ is referred to as an empirical distribution of $\boldsymbol{q}_s^{od}$ based on N^Q i.i.d. samples. The mth entry records the proportion of time that the sample path spends in state m. $\mathcal{P}(\varXi)$ is the set of all distributions supported on $\varXi$. $\boldsymbol{\varepsilon}$ is the radius of the Wasserstein ball centered at the empirical distribution $\hat{\mathbb{P}}_{N^Q}$.

E. Investment Risk Quantification

Proof 1: In fact, the decision variable for the optimization problem in this paper is $(\boldsymbol{x},\boldsymbol{y})$. For the sake of clarity, in this proof, $\boldsymbol{x}$ is used uniformly to represent the decision variable. It is assumed to belong to a set X, and $X \subseteq \mathbb{Z}^{o_1}\times\mathbb{R}^{o_2}$. Assuming that the dimension of the uncertain variable $\boldsymbol{q}$ is the same as that of $\boldsymbol{x}$, we set $o_0=o_1+o_2$ and define a set $\mathcal{O}=\{1,...,o_0\}$ indexed by o. In the problem context of this work, o can be equivalent to $s\times od$. Suppose that there is a finite data set $\hat{\mathcal{Q}}:=\{(\hat{q}_{o,1},...,\hat{q}_{o,N^Q})^\top, o\in\mathcal{O}\}$ for $\boldsymbol{q}$, where $\hat{q}_o$ belongs to a finite discrete support $\mathcal{Q}_O:=\{q_{o,1},...,q_{o,N^O}\}$ for some $q_{o,j}>0$ $\left(j\in\{1,...,N^O\}\right)$. And it is assumed that every element in $\hat{\mathcal{Q}}$ has its individual support, hence, there is no correlation information between components.

According to the empirical marginal distributions in the ambiguity set, $\hat{p}_{o,j}=\frac{1}{N^Q}\sum_{n=1}^{N^Q}\mathbb{1}(\hat{q}_{o,n}=q_{o,j})$ for $\forall j\in\{1,...,N^O\}$. We set the cost function corresponding to $\hat{q}_{o,n}$ as $\hat{c}_{o,n}$. Let $\overline{c}_o=\max\limits_{j\in\{1,...,N^O\}}c_{o,j}$ and $\underline{c}_o=\min\limits_{j\in\{1,...,N^O\}}c_{o,j}$ for each $o\in\mathcal{O}$, which are the upper and lower bounds of cost vector, respectively. And define a positive bias δ_o, which can be regarded as a fluctuating error.

Therefore, we can define $\hat{c}(\boldsymbol{x},\hat{\mathbb{P}}_{N^Q},N^Q)=\sum_{o\in\mathcal{O}}\min\{\frac{1}{N^Q}\sum_{n=1}^{N^Q}\overline{c}_{o,n}+\delta_o;\overline{c}_o\}$. As the true probability distribution $\mathbb{P}$ of uncertainty variable is unknown, an ideal decision would have to minimize the out-of-sample risk simultaneously for all $\mathbb{P}\in\mathcal{P}$. Hence, the probability part of out-of-sample disappointment constraint can be written as

$$\begin{aligned}
&\mathbb{P}^\infty\left(c(\boldsymbol{x},\mathbb{P})>\hat{c}\left(\boldsymbol{x},\hat{\mathbb{P}}_{N^Q},N^Q\right)\right)\\
&=\Pr\left\{\sum_{o\in\mathcal{O}}\left(\mathbb{E}_\mathbb{P}\{c_o\}-\min\left\{\frac{1}{N^Q}\sum_{n=1}^{N^Q}\hat{c}_{o,n}+\delta_o;\overline{c}_o\right\}\right)>0\right\}\\
&\leqslant\Pr\left\{\bigvee_{o\in\mathcal{O}}\left(\mathbb{E}_\mathbb{P}\{c_o\}-\min\left\{\frac{1}{N^Q}\sum_{n=1}^{N^Q}\hat{c}_{o,n}+\delta_o;\overline{c}_o\right\}>0\right)\right\}\\
&\leqslant\sum_{o\in\mathcal{O}}\Pr\left\{\mathbb{E}_\mathbb{P}\{c_o\}>\min\left\{\frac{1}{N^Q}\sum_{n=1}^{N^Q}\hat{c}_{o,n}+\delta_o;\overline{c}_o\right\}\right\}\\
&\leqslant\sum_{o\in\mathcal{O}}\Pr\left\{\left(\mathbb{E}_\mathbb{P}\{c_o\}>\frac{1}{N^Q}\sum_{n=1}^{N^Q}\hat{c}_{o,n}+\delta_o\right)\vee\left(\mathbb{E}_\mathbb{P}\{c_o\}>\overline{c}_o\right)\right\}\\
&\leqslant\sum_{o\in\mathcal{O}}\left(\Pr\left\{\mathbb{E}_\mathbb{P}\{c_o\}>\frac{1}{N^Q}\sum_{n=1}^{N^Q}\hat{c}_{o,n}+\delta_o\right\}+\Pr\left\{\mathbb{E}_\mathbb{P}\{c_o\}>\overline{c}_o\right\}\right)\\
&=\sum_{o\in\mathcal{O}}\Pr\left\{\mathbb{E}_\mathbb{P}\{c_o\}>\frac{1}{N^Q}\sum_{n=1}^{N^Q}\hat{c}_{o,n}+\delta_o\right\}\\
&=\sum_{o\in\mathcal{O}}\Pr\left\{\mathbb{E}_\mathbb{P}\{c_o\}-\frac{1}{N^Q}\sum_{n=1}^{N^Q}\hat{c}_{o,n}>\delta_o\right\}
\end{aligned}\tag{65}$$

Next, referring to [5], we set $\delta_o=(\overline{c}_o-\underline{c}_o)\sqrt{\frac{\varepsilon N^Q-\ln\delta}{2N^Q}}$, where the decay rate $\varepsilon>0$, parameter $\delta\in(0,1)$, and $\sum_{o\in\mathcal{O}}\delta=1$. Applying the Hoeffding's inequality in [6], we observe that

$$\begin{aligned}
&\Pr\left\{\mathbb{E}_\mathbb{P}\{c_o\}-\frac{1}{N^Q}\sum_{n=1}^{N^Q}\hat{c}_{o,n}>\delta_o\right\}\\
&\leqslant\left\{\left(\frac{\mu_1}{\mu_1+\mu_2}\right)^{\mu_1+\mu_2}\left(\frac{1-\mu_1}{1-\mu_1-\mu_2}\right)^{1-\mu_1-\mu_2}\right\}^{N^Q}\\
&\leqslant\exp\left[\frac{-2N^Q\delta_o^2}{\left(\overline{c}_o-\underline{c}_o\right)^2}\right]=\delta e^{-\varepsilon N^Q}
\end{aligned}\tag{66}$$

where $\mu_1=\frac{1}{N^Q}\sum_{n=1}^{N^Q}\hat{c}_{o,n}/\left(\overline{c}_o-\underline{c}_o\right)$ and $\mu_2=\delta_o/\left(\overline{c}_o-\underline{c}_o\right)$. Substituting the results of (66) into the original out-of-sample disappointment constraint, we obtain

$$\begin{aligned}
&\lim_{N^Q\to\infty}\frac{1}{N^Q}\log\mathbb{P}^\infty\left(c(\boldsymbol{x},\mathbb{P})>\hat{c}\left(\boldsymbol{x},\hat{\mathbb{P}}_{N^Q},N^Q\right)\right)\\
&\leqslant\lim_{N^Q\to\infty}\frac{1}{N^Q}\ln\left(\sum_{o\in\mathcal{O}}\delta e^{-\varepsilon N^Q}\right)\\
&\leqslant\lim_{N^Q\to\infty}\sup\frac{1}{N^Q}\ln e^{-\varepsilon N^Q}\\
&=-\varepsilon
\end{aligned}\tag{67}$$

In addition, we can obtain a conclusion that the optimization problem (29) with out-of-sample disappointment constraint satisfies the asymptotic performance and finite-sample guarantees [7]. □

Proof 2: Similarly as ***Proof 1***, we attempt to derive a tighter upper bound of out-of-sample disappointment by referring to the existing literature. We observe that the out-of-sample disappointment (29b) can be bounded as:

$$\begin{aligned}
&\mathbb{P}^\infty\left(c(\boldsymbol{x},\mathbb{P})>\hat{c}\left(\boldsymbol{x},\hat{\mathbb{P}}_{N^Q},N^Q\right)\right)\\
&=\Pr\left\{\sum_{o\in\mathcal{O}}\left(\mathbb{E}_\mathbb{P}\{c_o\}-\max_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_\mathbb{P}\{c_o\}\right)>0\right\}\\
&\leqslant\Pr\left\{\bigvee_{o\in\mathcal{O}}\left(\mathbb{E}_\mathbb{P}\{c_o\}-\max_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_\mathbb{P}\{c_o\}\right)>0\right\}\\
&\leqslant\sum_{o\in\mathcal{O}}\Pr\left\{\mathbb{E}_\mathbb{P}\{c_o\}-\max_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_\mathbb{P}\{c_o\}>0\right\}
\end{aligned}\tag{F_1}$$

Since the formula F_1 implies that there exist some sample distribution not in the empirical marginal distribution $\mathcal{P}$. Then it means that

$$\begin{aligned}
F_1&\leqslant\sum_{o\in\mathcal{O}}\Pr\{\mathbb{P}\notin\mathcal{P}\}\\
&=\sum_{o\in\mathcal{O}}\Pr\left\{\int_{\Xi^2}\|\boldsymbol{q}_o-\hat{\boldsymbol{q}}_o\|\Pi_o\left(d\boldsymbol{q}_o\times d\hat{\boldsymbol{q}}_o\right)\leqslant\varepsilon_o\right\}
\end{aligned}\tag{F_2}$$

Combining the measure concentration theory of Theorem 3.4 in [8], we can derive that

$$\begin{aligned}
F_2&=\sum_{o\in\mathcal{O}}\Pr\left\{d_o\left(\mathbb{P},\hat{\mathbb{P}}_{N^Q}\right)>\varepsilon_o\right\}\\
&\leqslant C_1\exp^{-C_2N^Q\varepsilon^{\max(2,\dim q)}}
\end{aligned}\tag{F_3}$$

for all $N^Q\geqslant1$, $\varepsilon_o>0$, and C_1 and C_2 are positive constants that only depend on the variable dimension and the number of samples.

Referring to the Proposition 8 and 10 in [9], we can obtain that

$$F_3 \leqslant \sum_{o\in\mathcal{O}} \sum_{j=1}^{N^O-1} \left(N^Q+1\right)^j \exp\left\{-N^Q \inf_{\hat{\mathbb{P}}_{N^Q}\in\mathcal{P}} d_o\left(\mathbb{P},\hat{\mathbb{P}}_{N^Q}\right)\right\} \tag{F_4}$$

Furthermore, the out-of-sample disappointment constraint can be transformed into

$$\begin{aligned}
&\lim_{N^Q\to\infty} \frac{1}{N^Q} \log \mathbb{P}^\infty \left(c\left(\boldsymbol{x},\mathbb{P}\right) > \hat{c}\left(\boldsymbol{x},\hat{\mathbb{P}}_{N^Q},N^Q\right)\right) \\
&\leqslant \lim_{N^Q\to\infty} \frac{1}{N^Q} \log\left(\sum_{o\in\mathcal{O}} \sum_{j=1}^{N^O-1} \left(N^Q+1\right)^j e^{-N^Q\varepsilon_o}\right) \\
&\leqslant \lim_{N^Q\to\infty} \frac{1}{N^Q} \log\left(\sum_{o\in\mathcal{O}} e^{-N^Q\varepsilon_o}\right) \\
&\leqslant \lim_{N^Q\to\infty} \frac{1}{N^Q} \sup \ln\left(e^{-N^Q\varepsilon}\right) \\
&= -\varepsilon
\end{aligned} \tag{71}$$

□

Remark 3: As the parameter N_s^{od} increases, the value on the right-hand side of the inequality (31) also increases. From this upper bound, an insightful preliminary conclusion can be drawn that the greater the number of possible values for discrete uncertain variables, the higher investment risk. This indicates that the degree of risk in this investment optimization problem is influenced not only by the fluctuation of uncertain variables but also by the number of discrete values in the support set.

REFERENCES

[1] M. V. Pereira, S. Granville, M. H. Fampa, R. Dix, L. A. Barroso, Strategic bidding under uncertainty: a binary expansion approach, IEEE Transactions on Power Systems 20 (1) (2005) 180–188.

[2] W. Gan, M. Shahidehpour, J. Guo, W. Yao, S. Pandey, E. A. Paaso, A. Vukojevic, J. Wen, A tri-level planning approach to resilient expansion and hardening of coupled power distribution and transportation systems, IEEE Transactions on Power Systems 37 (2) (2021) 1495–1507.

[3] J. Qiu, H. Yang, Z. Y. Dong, J. H. Zhao, K. Meng, F. J. Luo, K. P. Wong, A linear programming approach to expansion co-planning in gas and electricity markets, IEEE Transactions on Power Systems 31 (5) (2015) 3594–3606.

[4] K. S. Shehadeh, Distributionally robust optimization approaches for a stochastic mobile facility fleet sizing, routing, and scheduling problem, Transportation Science 57 (1) (2023) 197–229.

[5] B. P. Van Parys, P. M. Esfahani, D. Kuhn, From data to decisions: Distributionally robust optimization is optimal, Management Science 67 (6) (2021) 3387–3402.

[6] W. Hoeffding, Probability inequalities for sums of bounded random variables, The collected works of Wassily Hoeffding (1994) 409–426.

[7] T. Sutter, B. P. Van Parys, D. Kuhn, A pareto dominance principle for data-driven optimization, Operations Research (2024).

[8] P. Mohajerin Esfahani, D. Kuhn, Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations, Mathematical Programming 171 (1) (2018) 115–166.

[9] N. Fournier, A. Guillin, On the rate of convergence in wasserstein distance of the empirical measure, Probability theory and related fields 162 (3) (2015) 707–738.