Context-Aware Cluster-Based Multi-Uncertainty-Set Distributionally Robust Chance-Constrained Optimal DC Power Flow

Qiaoyi Wang*, Wei Liu†, Feiyu Chen*, Bo Yu‡

March 25, 2026

Abstract

This paper proposes a context-aware multi-uncertainty-set distributionally robust chance-constrained DC optimal power flow model. Meteorological features are projected to partition the non-convex error support into a context-dependent decomposition of conditional local ambiguity sets, with conditional weights inferred via kernel regression. The minimax problem is reformulated into a finite-dimensional second-order cone program with proven asymptotic consistency. Out-of-sample tests on the IEEE 30-bus and RBTS systems demonstrate that the proposed model achieves a high level of empirical feasibility close to the prescribed reliability level and avoids excessive conservativeness induced by global convex approximations.

1 Introduction

The increasing penetration of renewable energy sources, particularly wind power, has significantly amplified uncertainty in power system operation. As a result, traditional deterministic optimal power flow (OPF) models are no longer sufficient to ensure reliable and cost-effective system performance. This has motivated extensive research on stochastic and robust OPF formulations that explicitly account for uncertainty [1, 2, 3, 4, 5, 6, 7, 8].

To address uncertainty, chance-constrained optimal power flow (CC-OPF) models have been widely adopted to control the probability of constraint violations [4, 5]. However, the effectiveness of CC-OPF relies heavily on accurate knowledge of the underlying probability distribution of uncertainties, which is typically unavailable in practice [6, 7]. To overcome this limitation, distributionally robust optimization (DRO) has been introduced into OPF, leading to distributionally robust chance-constrained OPF (DRCC-OPF) models [9, 10, 11, 12].

Among various DRO approaches, Wasserstein-based ambiguity sets have attracted significant attention due to their favorable theoretical properties and computational tractability [13, 14]. Most existing DRCC-OPF approaches construct ambiguity sets based on moment information or probability metrics such as the Wasserstein distance [13, 14]. However, these methods typically rely on a single global ambiguity set, implicitly assuming homogeneous uncertainty across different operating conditions. Such a global construction often leads to overly conservative decisions, as extreme scenarios are enforced even when they are unlikely under specific contexts.

*Chongqing Normal University, Chongqing National Center for Applied Mathematics, China. These authors contributed equally to this work.

†Beijing Normal University at Zhuhai, China. These authors contributed equally to this work.

‡Corresponding author: yubo@dlut.edu.cn. Chongqing Normal University.

In practice, renewable forecast errors are strongly dependent on contextual information, such as wind speed, direction, and temperature. Ignoring such context may result in inaccurate uncertainty representation and suboptimal operational decisions. This observation has motivated the development of context-aware DRO frameworks [15], where ambiguity sets are conditioned on observable features. However, existing contextual DRO models typically employ a single conditional ambiguity set, which is insufficient to capture the multimodal and regime-dependent structure of renewable uncertainty.

To address these limitations, we propose a cluster-based multi-uncertainty-set distributionally robust OPF framework that incorporates contextual information. Specifically, the uncertainty space is partitioned into multiple context-dependent regions, and a local ambiguity set is constructed for each region. This structure enables a more refined representation of uncertainty and avoids enforcing a global worst-case distribution across heterogeneous scenarios.

To further improve tractability, we exploit the aggregation of wind power forecast errors and reformulate the high-dimensional uncertainty into a one-dimensional representation. This allows us to derive a computationally tractable reformulation of the proposed model.

The main contributions of this paper are summarized as follows:

- We develop a context-aware multi-uncertainty-set DRCC-OPF model that captures the heterogeneous and multimodal structure of renewable uncertainty.
- We construct ambiguity sets based on clustered contextual information, improving the balance between robustness and operational cost.
- We derive a tractable reformulation of the proposed model via uncertainty aggregation and convex analysis.
- Numerical experiments on standard test systems demonstrate that the proposed approach achieves improved reliability and reduced conservativeness compared with existing methods.

The remainder of this paper is organized as follows. Section 2 presents the fundamental reformulation theorem. Section 3 provides the theoretical analysis. Section 4 introduces the DRCC-DC-OPF model and its deterministic reformulation. Section 5 summarizes the proposed context-aware cluster-based DRCC-DC-OPF workflow. Section **??** reports numerical results, and Section 6 concludes the paper.

2 Fundamental reformulation theorem

We first state the deterministic reformulation that underpins the proposed model. The function $f(x, \boldsymbol{\omega})$ represents a generic convex piecewise-linear functional, which will later be instantiated for both the cost function and the constraint violation function.

Theorem 1 (Deterministic reformulation under multi-uncertainty-set DRO)**.** *Assume that:*

C1. *The decision variable satisfies $x \in \mathcal{X}$, where $\mathcal{X}$ is a nonempty convex set. The loss function has the convex piecewise-linear form*

$$f(x, \boldsymbol{\omega}) = \max_{m \in \mathcal{M}} \{\boldsymbol{a}_m(x)^\top \boldsymbol{\omega} + b_m(x)\}, \tag{1}$$

where $\boldsymbol{a}_m(x) \in \mathbb{R}^n$ and $b_m(x) \in \mathbb{R}$. Assume that $f(x, \cdot)$ is Lipschitz continuous on each Ξ_k.

C2. *The global uncertainty set* $\Xi \subset \mathbb{R}^n$ *is representable as the union of finitely many nonempty, compact, and convex subsets* Ξ_k *with nonempty interiors,* $k \in \mathcal{K} := \{1, \ldots, K\}$*, namely,*

$$\Xi = \bigcup_{k=1}^{K} \Xi_k, \qquad \Xi_k \cap \Xi_\ell = \varnothing, \ k \neq \ell. \tag{2}$$

The empirical cluster weight is $\hat{w}_k = N_k/N$*, and the conditional empirical distribution is*

$$\hat{P}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} \delta_{\hat{\boldsymbol{\omega}}_{k,i}}. \tag{3}$$

C3. *The outer weight ambiguity set and the inner conditional distribution ambiguity sets are given by*

$$\mathcal{W} = \left\{ \boldsymbol{w} \in \mathbb{R}_+^K : \sum_{k=1}^{K} w_k = 1, \ \sum_{k=1}^{K} |w_k - \hat{w}_k| \leq \delta_w \right\}, \tag{4}$$

$$\mathcal{P}_k = \left\{ P_k \in \mathscr{P}(\Xi_k) : W_1(P_k, \hat{P}_k) \leq \delta_k \right\}, \qquad k = 1, \ldots, K. \tag{5}$$

The L_1 *ambiguity set is adopted due to its tractability and its equivalence to a robust counterpart of total variation distance.*

Then the two-level minimax DRO problem

$$\min_{x \in \mathcal{X}} \sup_{\boldsymbol{w} \in \mathcal{W}} \sum_{k=1}^{K} w_k \sup_{P_k \in \mathcal{P}_k} \mathbb{E}_{P_k}[f(x, \boldsymbol{\omega})] \tag{6}$$

is equivalent to the following finite-dimensional deterministic optimization problem:

$$\begin{aligned}
\min_{x,\Theta} \quad & \eta + \nu(\delta_w - 1) + \sum_{k=1}^{K} \hat{w}_k t_k \\
s.t. \quad & t_k + \eta - \nu \geq \lambda_k \delta_k + \frac{1}{N_k} \sum_{i=1}^{N_k} s_{ik}, && \forall k, \\
& s_{ik} \geq b_m(x) + S_{\Xi_k}(\boldsymbol{a}_m(x) + \boldsymbol{\gamma}_{ikm}) - \boldsymbol{\gamma}_{ikm}^\top \hat{\boldsymbol{\omega}}_{k,i}, && \forall k, i, m, \\
& \|\boldsymbol{\gamma}_{ikm}\|_* \leq \lambda_k, && \forall k, i, m, \\
& t_k \geq 0, \ \lambda_k \geq 0, \ \nu \geq 0, \ \eta \in \mathbb{R}, \ x \in \mathcal{X},
\end{aligned} \tag{7}$$

where $\Theta := \{\eta, \nu, \boldsymbol{t}, \boldsymbol{\lambda}, \boldsymbol{s}, \boldsymbol{\gamma}_{ikm}\}$ *and* $S_{\Xi_k}(\cdot)$ *denotes the support function of the local closed convex set* Ξ_k*. Since each* Ξ_k *is compact, the support function* S_{Ξ_k} *is finite-valued. Here,* ν *corresponds to the dual variable of the outer weight ambiguity set, while* $\boldsymbol{\gamma}_{ikm}$ *arises from the Wasserstein dual formulation.*

The detailed proof is deferred to A.

3 Convergence analysis

This section establishes asymptotic consistency for the empirical multi-uncertainty-set DRO backbone with sample-based reference weights. In particular, the contextual weighting mechanism introduced in Section 4 is not part of the asymptotic model. The context-dependent online weighting rule introduced later is an operational extension and is not covered by the present asymptotic result. We first introduce the assumptions and limit functionals, and then present the unconstrained and constrained consistency statements. All convergence statements in this section are understood in the almost sure sense. Since K is finite, all almost sure events can be intersected into a common full-probability event. For the purpose of asymptotic analysis, the partition $\{\Xi_k\}_{k=1}^K$ is assumed to be fixed.

3.1 Probability space and sampling setup

All random objects in this section are defined on a common probability space

$$(\Omega, \mathcal{F}, \mathbb{P}).$$

Let

$$\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots : \Omega \to \Xi$$

be an i.i.d. sequence of Ξ-valued random vectors with common law $P^* \in \mathscr{P}(\Xi)$. For each $N \in \mathbb{N}$, the empirical sample is

$$\mathcal{D}_N(\omega) := \{\boldsymbol{\omega}_1(\omega), \ldots, \boldsymbol{\omega}_N(\omega)\}.$$

All empirical quantities, including N_k, $\hat{w}_k^{(N)}$, $\hat{P}_k^{(N)}$, $\mathcal{W}_N$, $\mathcal{P}_{k,N}$, and $J_N(\cdot)$, are measurable functions of the sample $\mathcal{D}_N$. Throughout this section, all almost sure statements are understood with respect to the probability measure $\mathbb{P} on (\Omega, \mathcal{F})$.

3.2 Theoretical assumptions

Assumption 1 (Compactness)**.** *The decision space $\mathcal{X}$ and the uncertainty set Ξ are nonempty compact sets. Moreover, the uncertainty set admits a finite* measurable partition

$$\Xi = \bigcup_{k=1}^{K} \Xi_k,$$

where each $\Xi_k \subset \mathbb{R}^n$ is nonempty and compact, and

$$\Xi_k \cap \Xi_\ell = \varnothing, \quad \forall k \neq \ell.$$

Assumption 2 (Lipschitz continuity of the loss)**.** *There exists $L > 0$ such that for all $x, x' \in \mathcal{X}$ and all $\boldsymbol{\omega}, \boldsymbol{\omega}' \in \Xi$,*

$$|f(x, \boldsymbol{\omega}) - f(x', \boldsymbol{\omega}')| \leq L\big(\|x - x'\|_2 + \|\boldsymbol{\omega} - \boldsymbol{\omega}'\|_2\big). \tag{8}$$

In particular, for every fixed $x \in \mathcal{X}$, the map $f(x, \cdot)$ is L-Lipschitz on Ξ.

Assumption 3 (Boundedness)**.** *There exists $M_f > 0$ such that*

$$\sup_{x \in \mathcal{X}, \boldsymbol{\omega} \in \Xi} |f(x, \boldsymbol{\omega})| \leq M_f. \tag{9}$$

Assumption 4 (Sampling scheme)**.** *Assume that the samples* $\{\boldsymbol{\omega}_i\}_{i\geq 1}$ *are i.i.d. random vectors taking values in* Ξ*. For each sample size* $N \in \mathbb{N}$ *and each* $k \in \{1, \dots, K\}$*, define*

$$N_k := \sum_{i=1}^{N} \mathbf{1}_{\{\boldsymbol{\omega}_i \in \Xi_k\}}, \qquad \hat{w}_k^{(N)} := \frac{N_k}{N}.$$

Let

$$w_k^* := \mathbb{P}(\boldsymbol{\omega} \in \Xi_k), \qquad k = 1, \dots, K,$$

and assume that

$$w_k^* > 0, \qquad k = 1, \dots, K.$$

Assumption 5 (Vanishing ambiguity radii)**.** *The outer and inner ambiguity radii satisfy*

$$\delta_w(N) \to 0, \qquad \delta_k(N_k) \to 0. \tag{10}$$

3.3 Conditional ambiguity sets

For each $k \in \{1, \dots, K\}$, define the true conditional distribution

$$P_k^*(\cdot) := \mathbb{P}(\cdot \mid \boldsymbol{\omega} \in \Xi_k), \tag{11}$$

and, on the event $\{N_k \geq 1\}$, define the empirical conditional distribution

$$\hat{P}_k^{(N)} := \frac{1}{N_k} \sum_{i:\boldsymbol{\omega}_i \in \Xi_k} \delta_{\boldsymbol{\omega}_i}, \tag{12}$$

as well as the corresponding ambiguity set

$$\mathcal{P}_{k,N} := \left\{ P \in \mathscr{P}(\Xi_k) : W_1(P, \hat{P}_k^{(N)}) \leq \delta_k(N_k) \right\}. \tag{13}$$

Lemma 1 (Convergence of empirical weights)**.** *Under Assumption 4,*

$$\hat{w}_k^{(N)} \xrightarrow{a.s.} w_k^*, \quad k = 1, \dots, K.$$

Consequently,

$$\|\hat{\boldsymbol{w}}^{(N)} - \boldsymbol{w}^*\|_1 \xrightarrow{a.s.} 0, \qquad N_k \to \infty \ a.s.$$

Proof. The result follows from the strong law of large numbers; see, e.g., [16]. □

Lemma 2 (Conditional empirical measure convergence)**.** *Under Assumptions 1 and 4, for each* $k \in \{1, \dots, K\}$*, the empirical conditional distribution* $\hat{P}_k^{(N)}$ *converges weakly to* P_k^* *almost surely as* $N \to \infty$*.*

Proof. By Lemma 1, we have $N_k \to \infty$ almost surely. Let $\tau_k(1) < \tau_k(2) < \cdots$ denote the indices of the samples such that $\boldsymbol{\omega}_{\tau_k(j)} \in \Xi_k$. Then $\{\boldsymbol{\omega}_{\tau_k(j)}\}_{j\geq 1}$ is an i.i.d. sequence with common law P_k^*. Since $N_k \to \infty$ almost surely, the result follows from standard weak convergence results for empirical measures; see, e.g., [16, 17]. □

3.4 Limit functional definition

Under Assumption 5, the limit weight set and limit ambiguity sets collapse to singletons:

$$\mathcal{W}_\infty := \left\{ \boldsymbol{w} \in \mathbb{R}_+^K : \sum_{k=1}^K w_k = 1,\ \|\boldsymbol{w} - \boldsymbol{w}^*\|_1 \leq 0 \right\} = \{\boldsymbol{w}^*\}, \tag{14}$$

$$\mathcal{P}_k^\infty := \{P \in \mathscr{P}(\Xi_k) : W_1(P, P_k^*) \leq 0\} = \{P_k^*\}. \tag{15}$$

Define the empirical objective functional and the limit functional as

$$J_N(x) := \sup_{\boldsymbol{w} \in \mathcal{W}_N} \sum_{k=1}^K w_k \sup_{P_k \in \mathcal{P}_{k,N}} \mathbb{E}_{P_k}[f(x, \boldsymbol{\omega})], \tag{16}$$

$$J_\infty(x) = \sup_{\boldsymbol{w} \in \mathcal{W}_\infty} \sum_{k=1}^K w_k \sup_{P_k \in \mathcal{P}_k^\infty} \mathbb{E}_{P_k}[f(x, \boldsymbol{\omega})] = \sum_{k=1}^K w_k^* \mathbb{E}_{P_k^*}[f(x, \boldsymbol{\omega})], \tag{17}$$

3.5 Unconstrained consistency

Define the one-sided excess of A over B by

$$D(A, B) := \sup_{a \in A} \inf_{b \in B} \|a - b\|_2. \tag{18}$$

Let

$$\hat{v}_N := \min_{x \in \mathcal{X}} J_N(x), \qquad \hat{X}_N := \arg\min_{x \in \mathcal{X}} J_N(x), \tag{19}$$

$$v^* := \min_{x \in \mathcal{X}} J_\infty(x), \qquad X^* := \arg\min_{x \in \mathcal{X}} J_\infty(x). \tag{20}$$

Proposition 1 (Unconstrained consistency)**.** *Under Assumptions 1–5, the following hold almost surely as $N \to \infty$:*

(i) Uniform convergence of the objective functional:

$$\sup_{x \in \mathcal{X}} |J_N(x) - J_\infty(x)| \xrightarrow{a.s.} 0. \tag{21}$$

(ii) Single-sided outer convergence of the optimal solution set:

$$D(\hat{X}_N, X^*) \xrightarrow{a.s.} 0. \tag{22}$$

The detailed proof is deferred to B.

3.6 Constrained consistency

In addition to Assumptions 1–5, we impose the following conditions on the constraint functional.

Assumption 6 (Constraint convexity, continuity and boundedness)**.** *For any $\boldsymbol{\omega} \in \Xi$, the function $g(\cdot, \boldsymbol{\omega})$ is convex on $\mathcal{X}$. Moreover, g is continuous on $\mathcal{X} \times \Xi$, and there exist $L_g > 0$ and $M_g > 0$ such that for all $x, x' \in \mathcal{X}$ and all $\boldsymbol{\omega}, \boldsymbol{\omega}' \in \Xi$,*

$$|g(x, \boldsymbol{\omega}) - g(x', \boldsymbol{\omega}')| \leq L_g\big(\|x - x'\|_2 + \|\boldsymbol{\omega} - \boldsymbol{\omega}'\|_2\big), \tag{23}$$

$$\sup_{x \in \mathcal{X}, \boldsymbol{\omega} \in \Xi} |g(x, \boldsymbol{\omega})| \leq M_g. \tag{24}$$

In particular, for every fixed $x \in \mathcal{X}$, the map $g(x, \cdot)$ is L_g-Lipschitz on Ξ.

Assumption 7 (Slater feasibility)**.** *The limit constrained problem satisfies a Slater condition: there exists $x^\circ \in \mathcal{X}$ such that*

$$\Phi_\infty(x^\circ) < 0. \tag{25}$$

Define the empirical constrained problem by

$$\begin{aligned} v_N := \inf_{x \in \mathcal{X}} \quad & J_N(x) \\ \text{s.t.} \quad & \Phi_N(x) \le 0, \end{aligned} \tag{26}$$

where

$$\Phi_N(x) := \sup_{\boldsymbol{w} \in \mathcal{W}_N} \sum_{k=1}^{K} w_k \sup_{P_k \in \mathcal{P}_{k,N}} \mathbb{E}_{P_k}[g(x, \boldsymbol{\omega})]. \tag{27}$$

Let the empirical feasible-set mapping be

$$\Gamma_N := \{x \in \mathcal{X} : \Phi_N(x) \le 0\}. \tag{28}$$

Similarly, the limit constrained problem is

$$\begin{aligned} v_\infty := \inf_{x \in \mathcal{X}} \quad & J_\infty(x) \\ \text{s.t.} \quad & \Phi_\infty(x) \le 0, \end{aligned} \tag{29}$$

where

$$\Phi_\infty(x) := \sum_{k=1}^{K} w_k^* \mathbb{E}_{P_k^*}[g(x, \boldsymbol{\omega})]. \tag{30}$$

Define the limit feasible region by

$$\Gamma_\infty := \{x \in \mathcal{X} : \Phi_\infty(x) \le 0\}. \tag{31}$$

Proposition 2 (Constrained robust consistency)**.** *Under Assumptions 1–5 and 6–7, the following hold almost surely as $N \to \infty$:*

(i) Uniform convergence of the constraint functional:

$$\sup_{x \in \mathcal{X}} |\Phi_N(x) - \Phi_\infty(x)| \xrightarrow{a.s.} 0. \tag{32}$$

(ii) Painlevé–Kuratowski convergence of feasible sets [18]: *almost surely,*

$$\operatorname*{Limsup}_{N \to \infty} \Gamma_N \subset \Gamma_\infty, \qquad \Gamma_\infty \subset \operatorname*{Liminf}_{N \to \infty} \Gamma_N. \tag{33}$$

(iii) Optimal value convergence:

$$v_N \xrightarrow{a.s.} v_\infty. \tag{34}$$

(iv) Single-sided outer convergence of constrained minimizers: *if*

$$\hat{X}_N^c := \arg\min\{J_N(x) : x \in \Gamma_N\}, \qquad X_c^* := \arg\min\{J_\infty(x) : x \in \Gamma_\infty\}, \tag{35}$$

then

$$D(\hat{X}_N^c, X_c^*) \xrightarrow{a.s.} 0. \tag{36}$$

The detailed proof is deferred to C.

4 Data-driven contextual feature clustering and uncertainty-set construction

4.1 PCA-based feature subspace projection and meteorological mode mapping

When processing joint meteorological covariates for multiple wind farms, the raw feature tensor often suffers from the curse of dimensionality. In high-dimensional spaces, distance concentration phenomena may degrade clustering performance, and the empirical covariance matrices become unstable, which motivates dimensionality reduction prior to clustering. To address this problem, principal component analysis (PCA) [19] is applied to the standardized meteorological feature matrix $\tilde{Z}_{\text{raw}} \in \mathbb{R}^{N \times d_{\text{raw}}}$, extracting the leading d_z principal components to form an orthogonal projection matrix $V \in \mathbb{R}^{d_{\text{raw}} \times d_z}$. The low-dimensional contextual feature representation is

$$\boldsymbol{z} = V^\top \tilde{\boldsymbol{z}}_{\text{raw}}, \qquad \boldsymbol{z} \in \mathcal{Z} \subset \mathbb{R}^{d_z}, \qquad d_z \ll d_{\text{raw}}. \tag{37}$$

Given the joint observation dataset

$$\mathcal{D} = \{(\boldsymbol{z}_i, \boldsymbol{\omega}_i)\}_{i=1}^{N}, \tag{38}$$

K-means clustering [20] is performed in the reduced feature space:

$$\min_{\{\boldsymbol{c}_k\}_{k=1}^{K}} \sum_{k=1}^{K} \sum_{\boldsymbol{z}_i \in C_k} \|\boldsymbol{z}_i - \boldsymbol{c}_k\|_2^2, \tag{39}$$

where $\boldsymbol{c}_k$ is the centroid of cluster C_k. The corresponding local error-sample subset is

$$\Xi_k^{\text{data}} := \{\boldsymbol{\omega}_i : \ \boldsymbol{z}_i \in C_k\}, \tag{40}$$

and the conditional empirical distribution is

$$\hat{P}_k = \frac{1}{N_k} \sum_{\boldsymbol{\omega}_i \in \Xi_k^{\text{data}}} \delta_{\boldsymbol{\omega}_i}, \qquad N_k := |\Xi_k^{\text{data}}|. \tag{41}$$

4.2 Context-dependent decomposition of support sets

The global support of wind-power errors is generally nonconvex and multimodal. Directly replacing it by a single global convex hull may include physically implausible combinations and lead to excessive conservativeness. We therefore construct a context-dependent decomposition of the global ambiguity set into conditional local ambiguity sets:

$$\Xi = \bigcup_{k=1}^{K} \Xi_k, \qquad \Xi_k \cap \Xi_\ell = \varnothing. \tag{42}$$

Within each cluster, the empirical mean $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$ are estimated, defining a local confidence ellipsoid by

$$\Xi_k := \left\{ \boldsymbol{\mu}_k + \boldsymbol{\Sigma}_k^{1/2} \boldsymbol{u} : \ \|\boldsymbol{u}\|_2 \le \rho_k \right\}, \qquad k = 1, \dots, K. \tag{43}$$

This construction preserves the multimodal structure of the global uncertainty support while maintaining the convexity required by the dual reformulation. This decomposition can be interpreted as a convexification-by-partition strategy, which preserves the intrinsic multimodal structure of the uncertainty while avoiding the excessive conservativeness induced by global convex hull approximations.

4.3 Dynamic inference of reference weights

At the online dispatch stage, the target-day forecast feature $\boldsymbol{z}_{\text{target}}$ is mapped to a dynamic contextual weight vector through a Softmax kernel regression mechanism:

$$\hat{w}_k(\boldsymbol{z}_{\text{target}}) = \frac{\exp\left(-\eta\|\boldsymbol{z}_{\text{target}} - \boldsymbol{c}_k\|_2^2\right)}{\sum_{j=1}^{K} \exp\left(-\eta\|\boldsymbol{z}_{\text{target}} - \boldsymbol{c}_j\|_2^2\right)}, \qquad k = 1, \dots, K, \tag{44}$$

where $\eta > 0$ controls the decay rate. This context-dependent weighting rule replaces the empirical frequency-based reference weights used in the asymptotic analysis. It is introduced to exploit forecast-side covariate information in online dispatch, and should be viewed as an operational extension of the empirical multi-uncertain-set DRO framework. The consistency results in Section 3 are established for the sample-based backbone model with empirical weights. When the target-day forecast is close to a certain weather regime centroid, the model places a larger reference weight on the corresponding local ambiguity set.

4.4 Affine-control DC-OPF architecture

Our model builds upon the contextual DRCC-OPF framework proposed in [15], while extending it to a cluster-based multi-uncertainty-set structure. Consider a power system with generator set $\mathcal{G}$, wind-farm set $\mathcal{W}$, bus set $\mathcal{B}$, and transmission-line set $\mathcal{C}$. The random uncertainty is the wind-power forecast error vector $\boldsymbol{\omega}$. Define the total forecast error by

$$\Omega := \mathbf{1}^\top \boldsymbol{\omega}. \tag{45}$$

This assumes a system-wide balancing policy, which is standard in affine reserve models. The day-ahead decision variables include the baseline generation g_j, downward and upward reserve capacities r_j^D, r_j^U, and participation factor β_j for each conventional generator $j \in \mathcal{G}$, as well as the load shedding ΔL_b for each bus $b \in \mathcal{B}$. In real time, the affine recourse policy is

$$\tilde{g}_j(\boldsymbol{\omega}) = g_j - \beta_j \Omega, \qquad \forall j \in \mathcal{G}. \tag{46}$$

The deterministic feasible set $\mathcal{X}$ of the decision vector $x = (g, \beta, r^D, r^U, \Delta L)$ contains the baseline physical constraints

$$g - r^D \geq g^{\min}, \qquad g + r^U \leq g^{\max}, \tag{47}$$

$$\langle \mathbf{1}, g\rangle + \langle \mathbf{1}, f\rangle = \langle \mathbf{1}, L - \Delta L\rangle, \qquad \langle \mathbf{1}, \beta\rangle = 1, \tag{48}$$

$$\beta \geq 0, \qquad r^D, r^U \geq 0, \qquad 0 \leq \Delta L \leq L. \tag{49}$$

In real time, the reserve and line-flow constraints become

$$-r^D \leq -\beta\Omega \leq r^U, \tag{50}$$

$$-\text{Cap} \leq M_G \tilde{g}(\boldsymbol{\omega}) + M_W(f + \boldsymbol{\omega}) - M_B(L - \Delta L) \leq \text{Cap}. \tag{51}$$

Here, M_G, M_W, and M_B denote the matrices of DC power transfer distribution factors (PTDFs) that map the power injections from generators, wind plants, and loads to the network line flows, respectively, while Cap represents the array of transmission line capacities [15]. These random inequalities must hold jointly with a probability of at least $1 - \epsilon$.

4.5 One-dimensional projection of the cost function

Under the affine recourse policy (46), the real-time output of each generator depends on the uncertainty only through the scalar quantity $\Omega = \mathbf{1}^\top \boldsymbol{\omega}$. Since the generation cost of each unit is approximated by a convex piecewise linear function of its output, the cost of each generator becomes a convex piecewise linear function of Ω. Consequently, the total operating cost can be expressed as:

$$C(\Omega) = \sum_{j \in \mathcal{G}} \max_{s=1,\dots,S_j} \{m_{js}(g_j - \beta_j \Omega) + n_{js}\}, \tag{52}$$

where S_j is the number of linear segments used for generator j, and m_{js} and n_{js} denote the slope and intercept of the s-th segment, respectively. Let $\Omega := \mathbf{1}^\top \boldsymbol{\omega}$ denote the aggregate wind power forecast error. Under this aggregation, the local ellipsoidal support set Ξ_k reduces to a one-dimensional interval

$$I_k = [\underline{\Omega}_k, \overline{\Omega}_k], \tag{53}$$

with bounds

$$\underline{\Omega}_k = \mathbf{1}^\top \boldsymbol{\mu}_k - \rho_k \sqrt{\mathbf{1}^\top \boldsymbol{\Sigma}_k \mathbf{1}}, \qquad \overline{\Omega}_k = \mathbf{1}^\top \boldsymbol{\mu}_k + \rho_k \sqrt{\mathbf{1}^\top \boldsymbol{\Sigma}_k \mathbf{1}}. \tag{54}$$

These bounds correspond to the extreme values of the ellipsoidal set under the linear projection $\Omega = \mathbf{1}^\top \boldsymbol{\omega}$. The dual robust expected-cost objective is therefore

$$\min_{x \in \mathcal{X}} \left\{ \langle c^D, r^D \rangle + \langle c^U, r^U \rangle + \langle c^{\text{shed}}, \Delta L \rangle + \sup_{\boldsymbol{w} \in \mathcal{W}(\boldsymbol{z}_{\text{target}})} \sum_{k=1}^{K} w_k \sup_{P_k^\Omega \in \mathcal{P}_k^\Omega} \mathbb{E}_{P_k^\Omega}[C(\Omega)] \right\}. \tag{55}$$

To facilitate the dual reformulation, we express the total cost in a unified max-affine form. For each generator $j \in \mathcal{G}$, define

$$f_{j,s}(\Omega) := m_{j,s}(g_j - \beta_j \Omega) + n_{j,s}, \quad s = 1, \dots, S_j.$$

Then

$$C(\Omega) = \sum_{j \in \mathcal{G}} \max_{s=1,\dots,S_j} f_{j,s}(\Omega).$$

Define the Cartesian product

$$\mathcal{M}_{\text{cost}} = \prod_{j \in \mathcal{G}} \{1, \dots, S_j\}.$$

Each element $(s_j)_{j \in \mathcal{G}} \in \mathcal{M}_{\text{cost}}$ represents a segment combination, where $s_j \in \{1, \dots, S_j\}$ specifies the selected segment of generator j.

Then the cost admits the equivalent max-affine representation

$$C(\Omega) = \max_{(s_j)_{j \in \mathcal{G}} \in \mathcal{M}_{\text{cost}}} \left\{ A_{(s_j)}(x)\, \Omega + B_{(s_j)}(x) \right\}, \tag{56}$$

where the coefficients are defined as

$$A_{(s_j)}(x) = -\sum_{j \in \mathcal{G}} m_{j,s_j} \beta_j, \qquad B_{(s_j)}(x) = \sum_{j \in \mathcal{G}} \big(m_{j,s_j} g_j + n_{j,s_j} \big). \tag{57}$$

The equivalence follows from the fact that the sum of generator-wise maxima can be represented as the maximum over all segment combinations. By applying Theorem 1 in the one-dimensional projected space, one obtains the finite-dimensional deterministic reformulation

$$\begin{aligned}\min_{x,\Theta^{\text{obj}}} \quad & \eta^{\text{obj}} + \nu^{\text{obj}}(\delta_w - 1) + \sum_{k=1}^{K} \hat{w}_k(\boldsymbol{z}_{\text{target}}) t_k^{\text{obj}} + \langle c^D, r^D \rangle + \langle c^U, r^U \rangle + \langle c^{\text{shed}}, \Delta L \rangle \\ \text{s.t.} \quad & t_k^{\text{obj}} + \eta^{\text{obj}} - \nu^{\text{obj}} \geq \lambda_k^{\text{obj}} \delta_k + \frac{1}{N_k} \sum_{i=1}^{N_k} s_{ik}^{\text{obj}}, \ \forall k, \\ & s_{ik}^{\text{obj}} \geq B_{(s_j)}(x) + S_{I_k}(A_{(s_j)}(x) + \gamma_{ik,(s_j)}^{\text{obj}}) - \gamma_{ik,(s_j)}^{\text{obj}} \hat{\Omega}_{k,i}, \ \forall k, i, (s_j)_{j \in \mathcal{G}} \in \mathcal{M}_{\text{cost}}, \\ & |\gamma_{ik,(s_j)}^{\text{obj}}| \leq \lambda_k^{\text{obj}}, \ \forall k, i, (s_j)_{j \in \mathcal{G}} \in \mathcal{M}_{\text{cost}}, \end{aligned} \tag{58}$$

where $\hat{\Omega}_{k,i} = \mathbf{1}^\top \hat{\boldsymbol{\omega}}_{k,i}$ and

$$S_{I_k}(v) = \max_{\Omega \in [\underline{\Omega}_k, \overline{\Omega}_k]} \{v\Omega\} = \max\{v\underline{\Omega}_k, v\overline{\Omega}_k\}. \tag{59}$$

4.6 CVaR approximation of the joint chance constraint

The random inequalities (50)–(51) can be unified as

$$\max_{m \in \mathcal{M}} \{\boldsymbol{a}_m(x)^\top \boldsymbol{\omega} + b_m(x)\} \leq 0. \tag{60}$$

Following Rockafellar and Uryasev [21], a standard conditional value-at-risk (CVaR) approximation of the corresponding joint chance constraint is

$$\inf_{\tau \in \mathbb{R}} \left\{ \tau + \frac{1}{\epsilon} \mathbb{E}_P \left[\left(\max_{m \in \mathcal{M}} \{\boldsymbol{a}_m(x)^\top \boldsymbol{\omega} + b_m(x)\} - \tau \right)^+ \right] \right\} \leq 0. \tag{61}$$

Using the law of total probability and the proposed weight–measure ambiguity structure, this yields

$$\inf_{\tau \in \mathbb{R}} \left\{ \tau + \frac{1}{\epsilon} \sup_{\boldsymbol{w} \in \mathcal{W}(\boldsymbol{z}_{\text{target}})} \sum_{k=1}^{K} w_k \sup_{P_k \in \mathcal{P}_k} \mathbb{E}_{P_k} \left[\left(\max_{m \in \mathcal{M}} \{\boldsymbol{a}_m(x)^\top \boldsymbol{\omega} + b_m(x)\} - \tau \right)^+ \right] \right\} \leq 0. \tag{62}$$

To absorb the positive-part operator, define the augmented index set $\mathcal{M}_+ := \mathcal{M} \cup \{0\}$. The auxiliary index $m = 0$ is introduced to absorb the positive-part operator in the CVaR reformulation. We set:

$$\tilde{\boldsymbol{a}}_m(x) = \begin{cases} \mathbf{0}, & m = 0, \\ \boldsymbol{a}_m(x), & m \in \mathcal{M}, \end{cases} \qquad \tilde{b}_m(x, \tau) = \begin{cases} 0, & m = 0, \\ b_m(x) - \tau, & m \in \mathcal{M}. \end{cases} \tag{63}$$

Then (62) becomes equivalent to the SOCP-representable system

$$\tau + \frac{1}{\epsilon} \left[\eta + \nu(\delta_w - 1) + \sum_{k=1}^{K} \hat{w}_k(\boldsymbol{z}_{\text{target}}) t_k \right] \leq 0, \tag{64}$$

$$t_k + \eta - \nu \geq \lambda_k \delta_k + \frac{1}{N_k} \sum_{i=1}^{N_k} s_{ik}, \qquad \forall k, \tag{65}$$

$$s_{ik} \geq \tilde{b}_m(x, \tau) + S_{\Xi_k}(\tilde{\boldsymbol{a}}_m(x) + \boldsymbol{\gamma}_{ikm}) - \boldsymbol{\gamma}_{ikm}^\top \hat{\boldsymbol{\omega}}_{k,i}, \qquad \forall k, i, m \in \mathcal{M}_+, \tag{66}$$

$$\|\boldsymbol{\gamma}_{ikm}\|_2 \leq \lambda_k, \qquad \forall k, i, m \in \mathcal{M}_+. \tag{67}$$

For the ellipsoidal support set (43), the support function has the analytical form

$$S_{\Xi_k}(\boldsymbol{v}) = \boldsymbol{v}^\top \boldsymbol{\mu}_k + \rho_k \|\boldsymbol{\Sigma}_k^{1/2} \boldsymbol{v}\|_2. \tag{68}$$

Substituting (68) into (66) yields a finite-dimensional SOCP reformulation. This reformulation preserves convexity and yields a tractable second-order cone program.

4.7 Deterministic equivalent master problem

Collecting the baseline constraints (47)–(49), the cost reformulation (58), and the SOCP chance-constraint reformulation (64)–(66), the complete deterministic equivalent model can be written as

$$\min_{\mathcal{X}_{\text{master}}} \quad \eta^{\text{obj}} + \nu^{\text{obj}}(\delta_w - 1) + \sum_{k=1}^{K} \hat{w}_k(\boldsymbol{z}_{\text{target}}) t_k^{\text{obj}} + \langle c^D, r^D \rangle + \langle c^U, r^U \rangle + \langle c^{\text{shed}}, \Delta L \rangle \tag{69a}$$

$$\text{s.t.} \quad g - r^D \geq g^{\min}, \qquad g + r^U \leq g^{\max}, \tag{69b}$$

$$\langle \mathbf{1}, g \rangle + \langle \mathbf{1}, f \rangle = \langle \mathbf{1}, L - \Delta L \rangle, \qquad \langle \mathbf{1}, \beta \rangle = 1, \tag{69c}$$

$$\tau + \frac{1}{\epsilon} \left[\eta + \nu(\delta_w - 1) + \sum_{k=1}^{K} \hat{w}_k(\boldsymbol{z}_{\text{target}}) t_k \right] \leq 0, \tag{69d}$$

$$t_k + \eta - \nu \geq \lambda_k \delta_k + \frac{1}{N_k} \sum_{i=1}^{N_k} s_{ik}, \qquad \forall k, \tag{69e}$$

$$s_{ik} \geq b_m(x) - \tau + (\boldsymbol{a}_m(x) + \boldsymbol{\gamma}_{ikm})^\top \boldsymbol{\mu}_k + \rho_k \|\boldsymbol{\Sigma}_k^{1/2} (\boldsymbol{a}_m(x) + \boldsymbol{\gamma}_{ikm})\|_2 - \boldsymbol{\gamma}_{ikm}^\top \hat{\boldsymbol{\omega}}_{k,i}, \quad \forall k, i, m \in \mathcal{M}, \tag{69f}$$

$$\|\boldsymbol{\gamma}_{ikm}\|_2 \leq \lambda_k, \qquad \forall k, i, m \in \mathcal{M}, \tag{69g}$$

$$s_{ik} \geq \boldsymbol{\gamma}_{ik0}^\top \boldsymbol{\mu}_k + \rho_k \|\boldsymbol{\Sigma}_k^{1/2} \boldsymbol{\gamma}_{ik0}\|_2 - \boldsymbol{\gamma}_{ik0}^\top \hat{\boldsymbol{\omega}}_{k,i}, \qquad \forall k, i, \tag{69h}$$

$$\|\boldsymbol{\gamma}_{ik0}\|_2 \leq \lambda_k, \qquad \forall k, i, \tag{69i}$$

$$t_k^{\text{obj}} + \eta^{\text{obj}} - \nu^{\text{obj}} \geq \lambda_k^{\text{obj}} \delta_k + \frac{1}{N_k} \sum_{i=1}^{N_k} s_{ik}^{\text{obj}}, \qquad \forall k, \tag{69j}$$

$$s_{ik}^{\text{obj}} \geq B_{(s_j)}(x) + \max\{(A_{(s_j)}(x) + \gamma_{ik,(s_j)}^{\text{obj}}) \underline{\Omega}_k, (A_{(s_j)}(x) + \gamma_{ik,(s_j)}^{\text{obj}}) \overline{\Omega}_k\} - \gamma_{ik,(s_j)}^{\text{obj}} \hat{\Omega}_{k,i}, \quad \forall k, i, (s_j)_{j \in \mathcal{G}} \in \mathcal{M}_{\text{cost}}, \tag{69k}$$

$$|\gamma_{ik,(s_j)}^{\text{obj}}| \leq \lambda_k^{\text{obj}}, \qquad \forall k, i, (s_j)_{j \in \mathcal{G}} \in \mathcal{M}_{\text{cost}}, \tag{69l}$$

$$t_k \geq 0, \ \lambda_k \geq 0, \ \nu \geq 0, \ \eta \in \mathbb{R}, \ \tau \in \mathbb{R}, \tag{69m}$$

$$t_k^{\text{obj}} \geq 0, \ \lambda_k^{\text{obj}} \geq 0, \ \nu^{\text{obj}} \geq 0, \ \eta^{\text{obj}} \in \mathbb{R}, \ \beta \geq 0, \ r^D, r^U \geq 0, \ 0 \leq \Delta L \leq L. \tag{69n}$$

The joint decision variable set is

$$\mathcal{X}_{\text{master}} = \{g, \beta, r^D, r^U, \Delta L, \tau, \eta, \nu, t, \lambda, s, \boldsymbol{\gamma}_{ikm}, \boldsymbol{\gamma}_{ik0}, \eta^{\text{obj}}, \nu^{\text{obj}}, t^{\text{obj}}, \lambda^{\text{obj}}, s^{\text{obj}}, \gamma_{ik,(s_j)}^{\text{obj}}\}. \tag{70}$$

The resulting problem is a second-order cone program that can be solved efficiently using off-the-shelf solvers such as MOSEK or Gurobi, with polynomial complexity in the sample size and number of clusters.

4.8 Explicit coefficient mappings

The random physical constraints are indexed by $\mathcal{M} = \{1, \ldots, 2|\mathcal{G}| + 2|\mathcal{C}|\}$. The coefficient vector $\boldsymbol{a}_m(x) \in \mathbb{R}^{|\mathcal{W}|}$ is defined by

$$\boldsymbol{a}_m(x) = \begin{cases} \beta_j \mathbf{1}, & m = j \quad \text{(downward reserve)}, \\ -\beta_j \mathbf{1}, & m = |\mathcal{G}| + j \quad \text{(upward reserve)}, \\ \left(M_{W,e,*} - (M_{G,e,*}\beta)\mathbf{1}^\top\right)^\top, & m = 2|\mathcal{G}| + e \quad \text{(forward line-flow violation)}, \\ -\left(M_{W,e,*} - (M_{G,e,*}\beta)\mathbf{1}^\top\right)^\top, & m = 2|\mathcal{G}| + |\mathcal{C}| + e \quad \text{(reverse line-flow violation)}. \end{cases} \tag{71}$$

The corresponding constant term $b_m(x)$ is

$$b_m(x) = \begin{cases} -r_j^D, & m = j, \\ -r_j^U, & m = |\mathcal{G}| + j, \\ M_{G,e,*}g + M_{W,e,*}f - M_{B,e,*}(L - \Delta L) - \mathrm{Cap}_e, & m = 2|\mathcal{G}| + e, \\ -\left(M_{G,e,*}g + M_{W,e,*}f - M_{B,e,*}(L - \Delta L)\right) - \mathrm{Cap}_e, & m = 2|\mathcal{G}| + |\mathcal{C}| + e. \end{cases} \tag{72}$$

Similarly, the piecewise cost coefficients are

$$A_{(s_j)}(x) = \sum_{j \in \mathcal{G}} -m_{j,s_j}\beta_j, \qquad B_{(s_j)}(x) = \sum_{j \in \mathcal{G}} (m_{j,s_j} g_j + n_{j,s_j}), \tag{73}$$

for $(s_j)_{j \in \mathcal{G}} \in \mathcal{M}_{\text{cost}}$.

Note that the cardinality of $\mathcal{M}_{\text{cost}}$ grows combinatorially with the number of generators; however, in practice, the number of segments per unit is small, which keeps the problem tractable.

5 Data-driven contextual dispatch algorithm

Algorithm 1 summarizes the proposed context-aware cluster-based DRCC-DC-OPF workflow.

Algorithm 1 Context-aware cluster-based multi-uncertainty-set DRCC-DC-OPF dispatch framework

Input: Historical joint observation dataset $\mathcal{D}_{\text{raw}} = \{(\tilde{\boldsymbol{z}}_{\text{raw},i}, \boldsymbol{\omega}_i)\}_{i=1}^{N}$; target-day raw meteorological forecast feature $\tilde{\boldsymbol{z}}_{\text{raw,target}}$; number of clusters K; Softmax decay parameter η; conservatism parameters ρ_k, δ_w, δ_k.

Output: Optimal dispatch and reserve decision x^*; dynamic contextual weights $\hat{w}$.

Offline stage: contextual partitioning
Compute the empirical mean $\boldsymbol{\mu}_Z$ and standard deviation $\boldsymbol{\sigma}_Z$ of the raw features, and standardize the data:
  $\tilde{\boldsymbol{z}}_i \leftarrow \text{diag}(\boldsymbol{\sigma}_Z)^{-1}(\tilde{\boldsymbol{z}}_{\text{raw},i} - \boldsymbol{\mu}_Z), \quad \forall i \in \{1, \ldots, N\}$.
Principal component analysis (PCA) [19] is applied to the standardized feature matrix, retaining the first d_z principal components to obtain the projection matrix V.
Map the raw contextual features to the reduced space according to (37):
  $\boldsymbol{z}_i \leftarrow V^\top \tilde{\boldsymbol{z}}_i, \quad \forall i \in \{1, \ldots, N\}$.
K-means clustering [20] is performed in the reduced feature space to obtain clusters $\{C_k\}_{k=1}^K$ and centroids $\{\boldsymbol{c}_k\}_{k=1}^K$.
for $k = 1$ to K **do**
  Build the local sample subset $\Xi_k^{\text{data}} := \{\boldsymbol{\omega}_i : \ \boldsymbol{z}_i \in C_k\}$ and estimate the local empirical mean $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$.
  Construct the local confidence ellipsoid Ξ_k as in (43).
end for
Online stage: contextual inference and dispatch
Standardize and project the target-day forecast feature to obtain $\boldsymbol{z}_{\text{target}}$.
Compute the reference weights $\hat{w}_k(\boldsymbol{z}_{\text{target}})$ via (44).
Substitute $\{\hat{w}_k(\boldsymbol{z}_{\text{target}}), \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}_{k=1}^K$ into the deterministic master problem (69).
Solve the resulting SOCP with a convex optimization solver.
Return the optimal decision x^*.

6 Conclusion

This paper proposed a context-aware cluster-based multi-uncertainty-set distributionally robust chance-constrained DC optimal power flow framework. By combining contextual feature clustering, a context-dependent decomposition of support sets, dynamic weight inference, and a weight–measure dual ambiguity structure, the proposed method captures multimodal weather-dependent wind uncertainty while retaining tractability. We established a finite-dimensional deterministic reformulation for both the objective and the CVaR approximation [21] of the joint chance constraints, and proved asymptotic consistency for the empirical DRO objective and constrained feasible-set mapping. Numerical experiments demonstrate that the proposed approach effectively mitigates global DRO over-approximation and achieves a better trade-off between expected operating cost and a high level of empirical feasibility close to the prescribed reliability level. Future work includes extending the framework to AC-OPF, integrating more expressive contextual clustering mechanisms, and conducting large-scale experiments on additional network benchmarks.

A Proof of Theorem 1

This section proves Theorem 1. The derivation consists of two steps: the dual reformulation of the outer weight simplex set and the Wasserstein dual reformulation of the inner conditional ambiguity sets.

A.1 Dual transformation of the outer weight simplex set

Let

$$v_k(x) := \sup_{P_k \in \mathcal{P}_k} \mathbb{E}_{P_k}[f(x, \boldsymbol{\omega})]. \tag{74}$$

Since $f(x, \cdot)$ is Lipschitz continuous on Ξ_k and Ξ_k is compact, $f(x, \cdot)$ is bounded on Ξ_k, and thus the expectation is well-defined.

Introduce nonnegative auxiliary variables $d_k^+, d_k^- \geq 0$ such that

$$w_k - \hat{w}_k = d_k^+ - d_k^-. \tag{75}$$

Then the outer worst-case expectation problem becomes

$$\begin{aligned} \sup_{\boldsymbol{w}, \boldsymbol{d}^+, \boldsymbol{d}^- \geq 0} \quad & \sum_{k=1}^K w_k v_k(x) \\ \text{s.t.} \quad & 1 - \sum_{k=1}^K w_k = 0, \\ & \hat{w}_k + d_k^+ - d_k^- - w_k = 0, \qquad \forall k, \\ & \delta_w - \sum_{k=1}^K (d_k^+ + d_k^-) \geq 0. \end{aligned} \tag{76}$$

Associate the Lagrange multipliers $\eta \in \mathbb{R}$, $\alpha_k \in \mathbb{R}$, and $\nu \geq 0$ with the above constraints. The Lagrangian function is

$$\begin{aligned} L = & \sum_{k=1}^K w_k v_k(x) + \eta\Big(1 - \sum_{k=1}^K w_k\Big) + \sum_{k=1}^K \alpha_k (\hat{w}_k + d_k^+ - d_k^- - w_k) \\ & + \nu\Big(\delta_w - \sum_{k=1}^K (d_k^+ + d_k^-)\Big) \\ = & \eta + \nu\delta_w + \sum_{k=1}^K \alpha_k \hat{w}_k + \sum_{k=1}^K w_k \big(v_k(x) - \eta - \alpha_k\big) + \sum_{k=1}^K d_k^+ (\alpha_k - \nu) + \sum_{k=1}^K d_k^- (-\alpha_k - \nu). \end{aligned} \tag{77}$$

By standard linear programming duality, the supremum of L over $w_k, d_k^+, d_k^- \geq 0$ is finite if and only if

$$v_k(x) - \eta - \alpha_k \leq 0, \qquad \alpha_k - \nu \leq 0, \qquad -\alpha_k - \nu \leq 0, \tag{78}$$

that is,

$$\alpha_k \in [v_k(x) - \eta, \nu] \cap [-\nu, \nu], \qquad \nu \geq 0. \tag{79}$$

Therefore,

$$\sup_{\boldsymbol{w} \in \mathcal{W}} \sum_{k=1}^K w_k v_k(x) = \inf_{\eta, \nu \geq 0, \boldsymbol{\alpha}} \left\{ \eta + \nu\delta_w + \sum_{k=1}^K \hat{w}_k \alpha_k : \alpha_k \geq v_k(x) - \eta, \ |\alpha_k| \leq \nu \right\}. \tag{80}$$

Since $\hat{w}_k \geq 0$, the optimal choice of α_k is the lower bound

$$\alpha_k = \max\{v_k(x) - \eta, -\nu\}. \tag{81}$$

Using $\max\{A,B\} = B + \max\{A-B,0\}$ and $\sum_{k=1}^K \hat{w}_k = 1$, we obtain

$$\begin{aligned}\sup_{\boldsymbol{w}\in\mathcal{W}} \sum_{k=1}^K w_k v_k(x) &= \inf_{\eta,\nu\geq 0} \left\{ \eta + \nu\delta_w + \sum_{k=1}^K \hat{w}_k \left[-\nu + \max\{v_k(x) - \eta + \nu, 0\}\right] \right\} \\ &= \inf_{\eta,\nu\geq 0} \left\{ \eta + \nu(\delta_w - 1) + \sum_{k=1}^K \hat{w}_k \max\{v_k(x) - \eta + \nu, 0\} \right\}. \end{aligned} \tag{82}$$

Introducing auxiliary variables $t_k \geq 0$ yields the equivalent representation

$$\begin{aligned} \inf_{\eta,\nu\geq 0,\, t_k\geq 0} \quad & \eta + \nu(\delta_w - 1) + \sum_{k=1}^K \hat{w}_k t_k \\ \text{s.t.} \quad & t_k + \eta - \nu \geq v_k(x), \qquad \forall k. \end{aligned} \tag{83}$$

A.2 Wasserstein dual reformulation of the inner conditional ambiguity set

The Wasserstein dual reformulation follows standard results in distributionally robust optimization (see, e.g., [13, 14]). Fix $k \in \{1,\dots,K\}$. Consider the inner worst-case expectation problem

$$v_k(x) = \sup_{P_k\in\mathcal{P}_k} \mathbb{E}_{P_k}[f(x,\boldsymbol{\omega})]. \tag{84}$$

By the definition of W_1, there exists a transportation plan $\Pi \in \mathscr{P}(\Xi_k \times \Xi_k)$ satisfying $\pi_{1\#}\Pi = P_k$ and $\pi_{2\#}\Pi = \hat{P}_k$. Since

$$\hat{P}_k = \frac{1}{N_k}\sum_{i=1}^{N_k} \delta_{\hat{\boldsymbol{\omega}}_{k,i}}, \tag{85}$$

measure disintegration implies the existence of conditional measures $\{\pi_i\}_{i=1}^{N_k} \subset \mathscr{P}(\Xi_k)$ such that

$$d\Pi(\boldsymbol{\omega},\hat{\boldsymbol{\omega}}) = \frac{1}{N_k}\sum_{i=1}^{N_k} d\pi_i(\boldsymbol{\omega})\,\delta_{\hat{\boldsymbol{\omega}}_{k,i}}(d\hat{\boldsymbol{\omega}}). \tag{86}$$

Thus,

$$\begin{aligned} v_k(x) = \sup_{\pi_i\in\mathscr{P}(\Xi_k)} \quad & \frac{1}{N_k}\sum_{i=1}^{N_k}\int_{\Xi_k} f(x,\boldsymbol{\omega})\,d\pi_i(\boldsymbol{\omega}) \\ \text{s.t.} \quad & \frac{1}{N_k}\sum_{i=1}^{N_k}\int_{\Xi_k} \|\boldsymbol{\omega} - \hat{\boldsymbol{\omega}}_{k,i}\|\,d\pi_i(\boldsymbol{\omega}) \leq \delta_k. \end{aligned} \tag{87}$$

Introduce the dual multiplier $\lambda_k \geq 0$ and form the Lagrangian

$$L(\pi_i,\lambda_k) = \lambda_k\delta_k + \frac{1}{N_k}\sum_{i=1}^{N_k}\int_{\Xi_k}\Big(f(x,\boldsymbol{\omega}) - \lambda_k\|\boldsymbol{\omega} - \hat{\boldsymbol{\omega}}_{k,i}\|\Big)d\pi_i(\boldsymbol{\omega}). \tag{88}$$

A Slater point exists because choosing $\tilde{\pi}_i = \delta_{\hat{\boldsymbol{\omega}}_{k,i}}$ yields transportation cost $0 < \delta_k$. Hence strong duality holds for this infinite-dimensional linear program; see, for example, the measure-linear-programming duality arguments used in Wasserstein DRO reformulations [13, 14]. Therefore,

$$\begin{aligned} v_k(x) &= \inf_{\lambda_k\geq 0}\sup_{\pi_i\in\mathscr{P}(\Xi_k)} L(\pi_i,\lambda_k) \\ &= \inf_{\lambda_k\geq 0}\left\{\lambda_k\delta_k + \frac{1}{N_k}\sum_{i=1}^{N_k}\sup_{\boldsymbol{\omega}\in\Xi_k}\left[f(x,\boldsymbol{\omega}) - \lambda_k\|\boldsymbol{\omega} - \hat{\boldsymbol{\omega}}_{k,i}\|\right]\right\}. \end{aligned} \tag{89}$$

Substituting (1) into (89), and introducing variables s_{ik}, we arrive at

$$t_k + \eta - \nu \geq \lambda_k \delta_k + \frac{1}{N_k} \sum_{i=1}^{N_k} s_{ik}, \tag{90}$$

with

$$s_{ik} \geq \sup_{\boldsymbol{\omega} \in \Xi_k} \{ f(x, \boldsymbol{\omega}) - \lambda_k \|\boldsymbol{\omega} - \hat{\boldsymbol{\omega}}_{k,i}\| \}. \tag{91}$$

Since

$$f(x, \boldsymbol{\omega}) = \max_{m \in \mathcal{M}} \{ \boldsymbol{a}_m(x)^\top \boldsymbol{\omega} + b_m(x) \}, \tag{92}$$

it suffices to impose, for all $m \in \mathcal{M}$,

$$s_{ik} - b_m(x) \geq \sup_{\boldsymbol{\omega} \in \Xi_k} \left\{ \boldsymbol{a}_m(x)^\top \boldsymbol{\omega} - \lambda_k \|\boldsymbol{\omega} - \hat{\boldsymbol{\omega}}_{k,i}\| \right\}. \tag{93}$$

Using the dual norm identity

$$-\lambda_k \|\boldsymbol{\omega} - \hat{\boldsymbol{\omega}}_{k,i}\| = \inf_{\|\boldsymbol{\gamma}_{ikm}\|_* \leq \lambda_k} \boldsymbol{\gamma}_{ikm}^\top (\boldsymbol{\omega} - \hat{\boldsymbol{\omega}}_{k,i}), \tag{94}$$

we can rewrite the right-hand side of (93) as

$$\sup_{\boldsymbol{\omega} \in \Xi_k} \inf_{\|\boldsymbol{\gamma}_{ikm}\|_* \leq \lambda_k} \left\{ \left(\boldsymbol{a}_m(x) + \boldsymbol{\gamma}_{ikm} \right)^\top \boldsymbol{\omega} - \boldsymbol{\gamma}_{ikm}^\top \hat{\boldsymbol{\omega}}_{k,i} \right\}. \tag{95}$$

Because the feasible set $\{\boldsymbol{\gamma}_{ikm} : \|\boldsymbol{\gamma}_{ikm}\|_* \leq \lambda_k\}$ is compact and convex, Ξ_k is compact and convex, and the integrand is bilinear, Sion's minimax theorem [22] allows us to exchange sup and inf:

$$\inf_{\|\boldsymbol{\gamma}_{ikm}\|_* \leq \lambda_k} \sup_{\boldsymbol{\omega} \in \Xi_k} \left\{ \left(\boldsymbol{a}_m(x) + \boldsymbol{\gamma}_{ikm} \right)^\top \boldsymbol{\omega} - \boldsymbol{\gamma}_{ikm}^\top \hat{\boldsymbol{\omega}}_{k,i} \right\}. \tag{96}$$

Invoking the support function

$$S_{\Xi_k}(\boldsymbol{v}) = \sup_{\boldsymbol{\omega} \in \Xi_k} \boldsymbol{v}^\top \boldsymbol{\omega}, \tag{97}$$

we obtain

$$s_{ik} \geq b_m(x) + S_{\Xi_k}(\boldsymbol{a}_m(x) + \boldsymbol{\gamma}_{ikm}) - \boldsymbol{\gamma}_{ikm}^\top \hat{\boldsymbol{\omega}}_{k,i}, \qquad \|\boldsymbol{\gamma}_{ikm}\|_* \leq \lambda_k. \tag{98}$$

Combining this with (83) completes the proof of Theorem 1. □

B Proof of Proposition 1

Proof. **Step 1: Pointwise convergence.** For any fixed $x \in \mathcal{X}$, by the triangle inequality,

$$|J_N(x) - J_\infty(x)| \leq A_N + B_N, \tag{99}$$

where

$$A_N := \left| \sup_{\boldsymbol{w} \in \mathcal{W}_N} \sum_{k=1}^{K} w_k \mathbb{E}_{\hat{P}_k^{(N)}}[f] - \sum_{k=1}^{K} w_k^* \mathbb{E}_{P_k^*}[f] \right|, \tag{100}$$

$$B_N := \left| \sup_{\boldsymbol{w} \in \mathcal{W}_N} \sum_{k=1}^{K} w_k \sup_{P_k \in \mathcal{P}_{k,N}} \mathbb{E}_{P_k}[f] - \sup_{\boldsymbol{w} \in \mathcal{W}_N} \sum_{k=1}^{K} w_k \mathbb{E}_{\hat{P}_k^{(N)}}[f] \right|. \tag{101}$$

Insert the intermediate term $\sup_{\boldsymbol{w}\in\mathcal{W}_N}\sum_k w_k \mathbb{E}_{P_k^*}[f]$ into A_N and use the subadditivity of the supremum:

$$\begin{aligned}
A_N &\le \left| \sup_{\boldsymbol{w}\in\mathcal{W}_N} \sum_{k=1}^{K} w_k \mathbb{E}_{\hat{P}_k^{(N)}}[f] - \sup_{\boldsymbol{w}\in\mathcal{W}_N} \sum_{k=1}^{K} w_k \mathbb{E}_{P_k^*}[f] \right| \\
&\quad + \left| \sup_{\boldsymbol{w}\in\mathcal{W}_N} \sum_{k=1}^{K} w_k \mathbb{E}_{P_k^*}[f] - \sum_{k=1}^{K} w_k^* \mathbb{E}_{P_k^*}[f] \right| \\
&\le \max_{1\le k\le K} \left| \mathbb{E}_{\hat{P}_k^{(N)}}[f] - \mathbb{E}_{P_k^*}[f] \right| + M_f \sup_{\boldsymbol{w}\in\mathcal{W}_N} \|\boldsymbol{w} - \boldsymbol{w}^*\|_1. \qquad (102)
\end{aligned}$$

For B_N, by the subadditivity of the supremum,

$$\begin{aligned}
B_N &\le \sup_{\boldsymbol{w}\in\mathcal{W}_N} \sum_{k=1}^{K} w_k \left(\sup_{P_k\in\mathcal{P}_{k,N}} \mathbb{E}_{P_k}[f] - \mathbb{E}_{\hat{P}_k^{(N)}}[f] \right) \\
&\le \max_{1\le k\le K} \sup_{P_k\in\mathcal{P}_{k,N}} \left(\mathbb{E}_{P_k}[f(x,\boldsymbol{\omega})] - \mathbb{E}_{\hat{P}_k^{(N)}}[f(x,\boldsymbol{\omega})] \right). \qquad (103)
\end{aligned}$$

Since $f(x,\cdot)$ is L-Lipschitz, the Kantorovich–Rubinstein duality yields

$$\left| \mathbb{E}_{P_k}[f(x,\boldsymbol{\omega})] - \mathbb{E}_{\hat{P}_k^{(N)}}[f(x,\boldsymbol{\omega})] \right| \le L\, W_1(P_k, \hat{P}_k^{(N)}). \qquad (104)$$

Taking supremum over $P_k \in \mathcal{P}_{k,N}$,

$$\sup_{P_k\in\mathcal{P}_{k,N}} \left| \mathbb{E}_{P_k}[f(x,\boldsymbol{\omega})] - \mathbb{E}_{\hat{P}_k^{(N)}}[f(x,\boldsymbol{\omega})] \right| \le L\, \delta_k(N_k). \qquad (105)$$

Hence,

$$B_N \le L \max_{1\le k\le K} \delta_k(N_k) \xrightarrow{a.s.} 0. \qquad (106)$$

Combining with the estimate for A_N,

$$|J_N(x) - J_\infty(x)| \xrightarrow{a.s.} 0. \qquad (107)$$

Step 2: Uniform convergence. For any $P_k \in \mathcal{P}_{k,N}$ and $x, x' \in \mathcal{X}$,

$$\left| \mathbb{E}_{P_k}[f(x,\boldsymbol{\omega})] - \mathbb{E}_{P_k}[f(x',\boldsymbol{\omega})] \right| \le L\|x - x'\|_2. \qquad (108)$$

Taking the supremum over P_k and $\boldsymbol{w} \in \mathcal{W}_N$,

$$|J_N(x) - J_N(x')| \le L\|x - x'\|_2, \qquad (109)$$

$$|J_\infty(x) - J_\infty(x')| \le L\|x - x'\|_2. \qquad (110)$$

Let $\varepsilon > 0$. $\exists \{x^1, \dots, x^M\} \subset \mathcal{X}$,

$$\mathcal{X} \subset \bigcup_{j=1}^{M} B\left(x^j, \frac{\varepsilon}{3L}\right). \qquad (111)$$

For every $x \in \mathcal{X}$, $\exists x^j$ with $\|x - x^j\|_2 \le \varepsilon/(3L)$.

$$\begin{aligned}
|J_N(x) - J_\infty(x)| &\le |J_N(x) - J_N(x^j)| + |J_N(x^j) - J_\infty(x^j)| + |J_\infty(x^j) - J_\infty(x)| \\
&\le \frac{2\varepsilon}{3} + \max_{1\le j\le M} |J_N(x^j) - J_\infty(x^j)|. \qquad (112)
\end{aligned}$$

Let $\{x^1, \ldots, x^M\} \subset \mathcal{X}$ be a finite δ-net with $\delta = \varepsilon/(3L)$, and define

$$\Omega_1 := \bigcap_{j=1}^{M} \left\{ \omega : \lim_{N\to\infty} J_N(x^j, \omega) = J_\infty(x^j) \right\}. \tag{113}$$

Then $\mathbb{P}(\Omega_1) = 1$. For any $\omega \in \Omega_1$, there exists $N_0(\omega)$ such that for all $N \geq N_0(\omega)$,

$$\max_{1\leq j\leq M} |J_N(x^j, \omega) - J_\infty(x^j)| \leq \varepsilon/3. \tag{114}$$

For any $x \in \mathcal{X}$, choose x^j with $\|x - x^j\|_2 \leq \delta$. By Lipschitz continuity,

$$|J_N(x, \omega) - J_N(x^j, \omega)| \leq \varepsilon/3, \qquad |J_\infty(x) - J_\infty(x^j)| \leq \varepsilon/3. \tag{115}$$

Thus,

$$\begin{aligned} |J_N(x, \omega) - J_\infty(x)| &\leq |J_N(x, \omega) - J_N(x^j, \omega)| \\ &\quad + |J_N(x^j, \omega) - J_\infty(x^j)| + |J_\infty(x^j) - J_\infty(x)| \\ &\leq \varepsilon. \end{aligned} \tag{116}$$

Hence,

$$\sup_{x\in\mathcal{X}} |J_N(x, \omega) - J_\infty(x)| \leq \varepsilon. \tag{117}$$

Step 4: Single-sided outer convergence of minimizers. Assume $\mathbb{P}(\limsup_{N\to\infty} D(\hat{X}_N, X^*) > 0) > 0$. $\exists \Omega_2, \mathbb{P}(\Omega_2) > 0$, $\forall \omega \in \Omega_2$, $\exists \varepsilon_0 > 0$, an increasing subsequence $\{N_m\}$, and $\hat{x}_{N_m}(\omega) \in \hat{X}_{N_m}(\omega)$,

$$\inf_{x^*\in X^*} \|\hat{x}_{N_m}(\omega) - x^*\|_2 \geq \varepsilon_0. \tag{118}$$

$\exists$ subsequence converging to $\tilde{x}(\omega) \in \mathcal{X}$.

$$\begin{aligned} |J_\infty(\tilde{x}(\omega)) - v^*| \leq |J_\infty(\tilde{x}(\omega)) - J_\infty(\hat{x}_{N_m}(\omega))| + |J_\infty(\hat{x}_{N_m}(\omega)) - J_{N_m}(\hat{x}_{N_m}(\omega), \omega)| \\ + |\hat{v}_{N_m}(\omega) - v^*|. \end{aligned} \tag{119}$$

For $\omega \in \Omega_1 \cap \Omega_2$, the limits yield $J_\infty(\tilde{x}(\omega)) = v^*$, $\tilde{x}(\omega) \in X^*$. Taking limits in (118) yields

$$\inf_{x^*\in X^*} \|\tilde{x}(\omega) - x^*\|_2 \geq \varepsilon_0, \tag{120}$$

contradicting $\tilde{x}(\omega) \in X^*$. Hence (22) holds. □

C Proof of Proposition 2

Proof. **Step 1: Uniform convergence of the constraint functional.** Since Assumptions 1, 4–5 are unchanged and Assumption 6 plays the role of Assumptions 2–3 for the constraint loss g, Proposition 1(i) applies verbatim with g replacing f. Consequently,

$$\sup_{x\in\mathcal{X}} |\Phi_N(x) - \Phi_\infty(x)| \xrightarrow{a.s.} 0.$$

Hence there exists an event Ω_3 with $\mathbb{P}(\Omega_3) = 1$ such that, for every $\omega \in \Omega_3$,

$$\lim_{N\to\infty} \sup_{x\in\mathcal{X}} |\Phi_N(x, \omega) - \Phi_\infty(x)| = 0.$$

Step 2: Eventual nonemptiness of the empirical feasible sets. $\exists x^\circ \in \mathcal{X}$, $\Phi_\infty(x^\circ) < 0$. Let $\eta_0 := -\frac{1}{2}\Phi_\infty(x^\circ) > 0$. $\forall \omega \in \Omega_3$, $\exists N_0(\omega) \in \mathbb{N}$, $\forall N \geq N_0(\omega)$,

$$\sup_{x\in\mathcal{X}} |\Phi_N(x,\omega) - \Phi_\infty(x)| \leq \eta_0. \tag{121}$$

$$\Phi_N(x^\circ,\omega) \leq \Phi_\infty(x^\circ) + \eta_0 < 0. \tag{122}$$

$x^\circ \in \Gamma_N(\omega)$, $\forall N \geq N_0(\omega)$. $\Gamma_N(\omega) \neq \varnothing$.

Step 3: Painlevé–Kuratowski convergence of feasible sets.

Outer inclusion. Let $x \in \operatorname{Limsup}_{N\to\infty} \Gamma_N(\omega)$. $\exists \{N_m\}$, $x_{N_m} \in \Gamma_{N_m}(\omega)$, $x_{N_m} \to x$.

$$\Phi_{N_m}(x_{N_m},\omega) \leq 0. \tag{123}$$

$$\begin{aligned}\Phi_\infty(x) &\leq |\Phi_\infty(x) - \Phi_\infty(x_{N_m})| + |\Phi_\infty(x_{N_m}) - \Phi_{N_m}(x_{N_m},\omega)| + \Phi_{N_m}(x_{N_m},\omega)\\ &\leq L_g\|x - x_{N_m}\|_2 + \sup_{z\in\mathcal{X}} |\Phi_\infty(z) - \Phi_{N_m}(z,\omega)|.\end{aligned} \tag{124}$$

Taking $m \to \infty$ yields $\Phi_\infty(x) \leq 0$, $x \in \Gamma_\infty$.

Inner inclusion. Let $x \in \Gamma_\infty$, $\Phi_\infty(x) \leq 0$. $\forall \lambda \in (0,1)$,

$$x^\lambda := \lambda x^\circ + (1-\lambda)x. \tag{125}$$

$$\Phi_\infty(x^\lambda) \leq \lambda\Phi_\infty(x^\circ) + (1-\lambda)\Phi_\infty(x) \leq \lambda\Phi_\infty(x^\circ). \tag{126}$$

Let $\eta_\lambda := -\lambda\Phi_\infty(x^\circ) > 0$. $\Phi_\infty(x^\lambda) \leq -\eta_\lambda$. $\forall \omega \in \Omega_3$, $\exists N_1(\omega) \in \mathbb{N}$, $\forall N \geq N_1(\omega)$,

$$\sup_{z\in\mathcal{X}} |\Phi_N(z,\omega) - \Phi_\infty(z)| \leq \eta_\lambda, \tag{127}$$

$$\Phi_N(x^\lambda,\omega) \leq \Phi_\infty(x^\lambda) + \eta_\lambda \leq 0. \tag{128}$$

$x^\lambda \in \Gamma_N(\omega)$, $\forall N \geq N_1(\omega)$.

$$\limsup_{N\to\infty} \inf_{y\in\Gamma_N(\omega)} \|x - y\|_2 \leq \limsup_{N\to\infty} \|x - x^\lambda\|_2 = \lambda\|x - x^\circ\|_2. \tag{129}$$

Letting $\lambda \downarrow 0$ yields $x \in \operatorname{Liminf}_{N\to\infty} \Gamma_N(\omega)$.

Step 4: Constrained optimal value convergence. Let $x^* \in X_c^* \subset \Gamma_\infty$. $\Gamma_\infty \subset \operatorname{Liminf} \Gamma_N(\omega)$, $\exists x_N \in \Gamma_N(\omega)$, $x_N \to x^*$.

$$v_N(\omega) = \inf_{x\in\Gamma_N(\omega)} J_N(x,\omega) \leq J_N(x_N,\omega). \tag{130}$$

$$\limsup_{N\to\infty} v_N(\omega) \leq \lim_{N\to\infty} J_N(x_N,\omega) = J_\infty(x^*) = v_\infty. \tag{131}$$

$\exists \{N_m\}$,

$$\lim_{m\to\infty} v_{N_m}(\omega) = \liminf_{N\to\infty} v_N(\omega). \tag{132}$$

Let $\hat{x}_{N_m}(\omega) \in \hat{X}^c_{N_m}(\omega) \subset \Gamma_{N_m}(\omega)$. $\exists$ subsequence, $\hat{x}_{N_m}(\omega) \to \bar{x}(\omega) \in \mathcal{X}$. $\bar{x}(\omega) \in \Gamma_\infty$.

$$\begin{aligned}|v_{N_m}(\omega) - J_\infty(\bar{x}(\omega))| &= |J_{N_m}(\hat{x}_{N_m}(\omega),\omega) - J_\infty(\bar{x}(\omega))|\\ &\leq \sup_{z\in\mathcal{X}} |J_{N_m}(z,\omega) - J_\infty(z)| + L\|\hat{x}_{N_m}(\omega) - \bar{x}(\omega)\|_2.\end{aligned} \tag{133}$$

$\lim_{m\to\infty} |v_{N_m}(\omega) - J_\infty(\bar{x}(\omega))| = 0$.

$$\liminf_{N\to\infty} v_N(\omega) = J_\infty(\bar{x}(\omega)) \geq \min_{x\in\Gamma_\infty} J_\infty(x) = v_\infty. \tag{134}$$

$$v_N \xrightarrow{a.s.} v_\infty \tag{135}$$

Step 5: Single-sided outer convergence of constrained minimizers. Assume $\mathbb{P}(\limsup_{N\to\infty} D(\hat{X}_N^c, X_c^*) > 0) > 0$. $\exists\Omega_4, \mathbb{P}(\Omega_4) > 0$, $\forall\omega \in \Omega_4$, $\exists\varepsilon_0 > 0$, an increasing subsequence $\{N_m\}$, $\hat{x}_{N_m}(\omega) \in \hat{X}_{N_m}^c(\omega)$,

$$\inf_{x^*\in X_c^*} \|\hat{x}_{N_m}(\omega) - x^*\|_2 \geq \varepsilon_0. \tag{136}$$

$\exists$ subsequence, $\hat{x}_{N_m}(\omega) \to \tilde{x}(\omega) \in \mathcal{X}$. $\tilde{x}(\omega) \in \Gamma_\infty$.

$$\begin{aligned}|J_\infty(\tilde{x}(\omega)) - v_\infty| \leq\ & |J_\infty(\tilde{x}(\omega)) - J_\infty(\hat{x}_{N_m}(\omega))| + |J_\infty(\hat{x}_{N_m}(\omega)) - J_{N_m}(\hat{x}_{N_m}(\omega), \omega)| \\ & + |v_{N_m}(\omega) - v_\infty|.\end{aligned} \tag{137}$$

For $\omega \in \Omega_3 \cap \Omega_4$, the limits yield $J_\infty(\tilde{x}(\omega)) = v_\infty$, $\tilde{x}(\omega) \in X_c^*$. Taking limits in the distance inequality yields

$$\inf_{x^*\in X_c^*} \|\tilde{x}(\omega) - x^*\|_2 \geq \varepsilon_0, \tag{138}$$

contradicting $\tilde{x}(\omega) \in X_c^*$. Hence (36) holds. □

References

[1] A. L. Soyster, Convex programming with set-inclusive constraints and applications to inexact linear programming, Operations Research 21 (5) (1973) 1154–1157.

[2] A. Ben-Tal, L. El Ghaoui, A. Nemirovski, Robust Optimization, Princeton University Press, Princeton, 2009.

[3] J. R. Birge, F. Louveaux, Introduction to Stochastic Programming, 2nd Edition, Springer, New York, 2011.

[4] P. Zhang, F. Li, Chance constrained programming for optimal power flow under uncertainty, IEEE Transactions on Power Systems 26 (4) (2011) 2417–2424.

[5] M. Vrakopoulou, K. Margellos, J. Lygeros, G. Andersson, Chance constrained reserve scheduling using uncertain wind power forecasts, IFAC Proceedings Volumes 46 (27) (2013) 71–76.

[6] M. Lubin, Y. Dvorkin, S. Backhaus, Robust and chance-constrained optimal power flow, IEEE Transactions on Power Systems 31 (5) (2016) 3847–3856.

[7] A. Peña-Ordieres, D. K. Molzahn, L. Roald, A. Wächter, Improved formulations for joint chance-constrained ac optimal power flow, Electric Power Systems Research 186 (2020) 106349.

[8] Y. Jia, G. Hug, S. Shen, Chance-constrained economic dispatch with joint constraints, IEEE Transactions on Power Systems 36 (2) (2021) 1206–1217.

[9] E. Delage, Y. Ye, Distributionally robust optimization under moment uncertainty with application to data-driven problems, Operations Research 58 (3) (2010) 595–612.

[10] Z. Wang, P. Glynn, Y. Ye, Likelihood robust optimization for data-driven problems, Computational Management Science 13 (2) (2016) 241–261.

[11] H. Rahimian, S. Mehrotra, Distributionally robust optimization: A review, arXiv preprint arXiv:1908.05659 (2022).

[12] D. Kuhn, S. Shafiee, W. Wiesemann, Distributionally robust optimization, arXiv preprint arXiv:2411.02549 (2024).

[13] P. Mohajerin Esfahani, D. Kuhn, Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations, Mathematical Programming 171 (1–2) (2018) 115–166.

[14] J. Blanchet, K. Murthy, Quantifying distributional model risk via optimal transport, Mathematics of Operations Research 44 (2) (2019) 565–600.

[15] A. Esteban-Pérez, J. M. Morales, Distributionally robust optimal power flow with contextual information, European Journal of Operational Research 306 (3) (2023) 1047–1058.

[16] P. Billingsley, Probability and Measure, 3rd Edition, Wiley, 1995.

[17] A. W. van der Vaart, J. A. Wellner, Weak Convergence and Empirical Processes, Springer, 1996.

[18] R. T. Rockafellar, R. J.-B. Wets, Variational Analysis, Springer, Berlin, 1998.

[19] I. T. Jolliffe, Principal Component Analysis, 2nd Edition, Springer, New York, 2002.

[20] S. Lloyd, Least squares quantization in pcm, IEEE Transactions on Information Theory 28 (2) (1982) 129–137.

[21] R. T. Rockafellar, S. Uryasev, Optimization of conditional value-at-risk, Journal of Risk 2 (3) (2000) 21–42.

[22] M. Sion, On general minimax theorems, Pacific Journal of Mathematics 8 (1) (1958) 171–176.