Minimax Theory and its Applications
Volume 11 (2026), No. 4s, 837 to 847

Robust Optimization of Learning Models Under Distributional and Parametric Uncertainty

Dr. Gururaja H S[1], Dr. S. Swarnalatha[2], Anto Nibiya S[3], Dr. Ramesh N S V S C Sripada[4], Kakkireni Deepthi[5], S. Jeyanthi[6]

[1]Associate Professor, Department of Computer Science and Business Systems, B.M.S. College of Engineering, Bengaluru, Email: gururajhs.ise@bmsce.ac.in, ORCID ID: 0000-0002-9718-4672
[2]Associate Professor, Department of Hindi, Sri Ramakrishna College of Arts & Science, Email: swarnalathar023@gmail.com, ORCID ID: 0009-0005-3061-8623
[3]Assistant Professor, Department of Mathematics, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Email: antonibiya@maher.ac.in, ORCID ID: 0009-0004-1888-792X
[4]Associate Professor, Department of Artificial Intelligence and Machine Learning, Aditya University, Surampalem, Andhra Pradesh – 533437, Email: rameshnsvscs@adityauniversity.in, ORCID ID: 0000-0003-3507-5363
[5]Assistant Professor, Dept. of ECE, sreenidhi Institute of science and Technology, Email: deepthi.k@sreenidhi.edu.in, orcid id: https://orcid.org/0009-0004-7727-0575
[6]Department of Computer Applications, New Prince Shri Bhavani College of Engineering and Technology
Email: prabakaran.jeyanthi@gmail.com, ORCID ID: 0000-0002-4003-8658

Abstract
Learning models in uncertain environments are subjected to many uncertainties, such as uncertainty in the data distribution and uncertainty in the model parameters. These uncertainty sources are usually handled separately in conventional robust learning methods, or with fixed robustness levels, causing robustness to be allocated ineffectively and introducing conservatism. In this work, a sensitivity-adaptive robust learning approach that jointly accounts for both distributional and parametric uncertainty with a shared robustness budget is proposed. Distributional uncertainty is described by a Wasserstein ambiguity set, whereas parametric uncertainty is described by an ellipsoidal uncertainty set. The proposed framework quantifies the relative sensitivity of the learning objective with respect to these uncertainty sources and allocates the robustness budget in response to the changing uncertainty sensitivity profile. A mathematical formulation is provided to explain the integrated robust-learning objective, sensitivity-based allocation mechanism, joint worst-case risk and projected adaptive update. Compared with empirical risk minimization, Wasserstein distributionally robust optimization, parametric robust learning and fixed coupled robustness, a numerical analysis is carried out. Results show that the sensitivity-adaptive allocation can minimize robust risk while providing a more balanced treatment of distributional uncertainty and parametric uncertainty. The allocation is also steadily convergent, with the sensitivity profile remaining steady. The study offers a coherent framework for handling several uncertainty sources in learning models and thus a basis for adaptive robustness in uncertain learning contexts.

Keywords: Robust Learning, Distributional Uncertainty, Parametric Uncertainty, Wasserstein Distributionally Robust Optimization, Sensitivity-Adaptive Allocation, Robustness Budget, Uncertainty-Aware Learning.

1. Introduction

The growing use of learning models in uncertain and changing environments has led to the need for approaches that can be relied upon to perform well when operating outside conditions simulated by training data. In a traditional approach to learning, it is often assumed that past statistics of observed samples are representative of the future. During deployment, however, distributional shift, heterogeneity of operating environments, measurement fluctuation, and uncertainty of model parameters may decrease learning performance. Therefore, it is important that models learn strongly when having to work in environments with multiple different uncertainties. Distributionally robust optimization (DRO) and robust optimization offer tools to deal with uncertainty in the probability distribution and in the model parameters, respectively.

A significant drawback of Empirical Risk Minimization (ERM) is that it is based on the observed training distribution. ERM minimizes average loss, given seen data points, but does not directly prevent underperformance with respect to differences in training and deployment distributions. The motivation for Wasserstein DRO is to

remedy this by taking into account plausible probability distributions around the empirical distribution and optimize under adverse distributional conditions. The recent study has drawn parallels between Wasserstein DRO, generalization, and regularization, rendering it more theoretically significant to learning tasks (Wu et al., 2025; Gao et al., 2024). Recent innovations such as duality and algorithms for Wasserstein-based optimization (Zhang et al., 2024; Wang et al., 2025) have further enhanced the applicability of Wasserstein-based optimization.

The spread of the distribution is not the only source of instability. Model parameters can also be uncertain due to estimation errors, approximation, imperfect model calibration and/or changes in the underlying system. Robust optimization offers a methodology to deal with this type of uncertainty by encoding uncertain parameters with specified uncertainty sets. With the development of adjustable robust optimisation, decisions are known to be very important in uncertain situations (Wei & Zhang, 2024; Detienne et al., 2024). The research on uncertainty coupling also shows that the simultaneous analysis of multiple uncertainty sources can offer potential benefits over the method of analyzing uncertainties individually. (Wang et al., 2024)

An important challenge that has yet to be settled is the allocation of robustness in cases where both types of uncertainty (distributional and parametric) are present. Current methods typically give their robustness values separately with no regard for the nature of the learning objective: whether it is more susceptible to one or another source of robustness. This independent specification may lead to an inefficient allocation of robustness or to being unnecessarily conservative. Sensitivity-based interaction between distributional and parametric robustness is not well developed in the learning setting considered here, although other research has progressed in Wasserstein DRO, robust learning and adaptive optimization.

This paper presents a sensitivity-adaptive robust learning framework to compensate for such uncertainty, and combines both distributional and parametric uncertainty, subject to a joint robustness budget. The framework involves a comparison of the sensitivity of the learning loss to each source of uncertainty and a calibration of the robustness transfers to the other. The philosophy of the contribution is not the introduction of Wasserstein DRO per se, but a coupling mechanism based on sensitivity for coordinating both distributional and parametric protection. Numerically controlled quasi-experimental analysis of the convergence of adaptive allocation is also supported by theoretical characterization of joint robust risk behaviour.

The study has three key contributions. It proposes, first, a coherent approach to simultaneous distributional and parametric uncertainty in learning. Secondly, it provides a sensitivity-based way of distributing a finite robustness resource among these uncertainty sources. Third, the resulting adaptive robustness behaviour is theoretically and numerically analyzed. These contributions combine to create an interpretable method to coordinate uncertainty protection based on learning-model sensitivity

2. Literature Review

2.1 Wasserstein Distributionally Robust Optimization

In this section,we revisit Wasserstein Distributionally Robust Optimization (WRDO). Wasserstein distributionally robust optimization (DRO) offers a systematic method to deal with uncertainty in the distribution of observed data. It deals with plausible distributions within a Wasserstein distance to the empirical distribution and optimizes the performance against adverse distributions. This method has been well grounded with theoretical support and offers finite-sample guarantees when optimizing based on data. It has recently also been shown to be relevant for strong learning and distributional uncertainty

2.2 Generalization and Regularization

Wasserstein DRO relates to generalization and regularization of learning models. Distributional robustness can impact model behaviour and provide greater model certainty in shifting data conditions by controlling deviation from the empirical distribution. Research exists that has associated Wasserstein robustness with variation regularization and generalization; a framework for establishing learning models that perform well outside the observed training distribution

2.3 Wasserstein Duality and Computational Reformulations

A significant problem in Wasserstein DRO is minimizing over probability distributions. Thus, duality-based formulations and computational reformulations have been extensively studied. These advances offer convenient representations of distributionally robust problems as well as enhance their usability in realistic learning and optimization scenarios. Moreover, some efficient methods have been proposed to alleviate the computational burden of the Wasserstein-based formulations.

2.4 Adaptive and Adjustable Robust Optimization

Adaptive and adjustable robust optimization is concerned with adapting decisions to different uncertainty scenarios and not necessarily to static configurations. This understanding is especially important in the case when the uncertainty characteristics change during optimization. Previous studies have shown that decisions that can be varied after design can help enhance the flexibility of robust optimization and give more response to nuclear objectives and nuclear operating conditions when they are subjected to uncertainty.

3. Uncertainty Modelling

3.1 Empirical Distribution

The proposed framework starts from the empirical description of the learning environment, which is captured from observations of the available training data. The empirical distribution is used as the reference in considering the possibility of changes in the data-generating environment. In traditional learning, this distribution is generally considered a reasonably good approximation of the distribution of the underlying population. But this assumption can be limiting if all the conditions represented in the training observations are not considered in deployment conditions.

The empirical distribution is consequently being used in the proposed framework not as a representation of the future operating environment but rather as a reference. Observed training samples are used to determine the information that is needed for modelling these observed training samples as a reference to the current learning problem, and uncertainty models reflect the fact that future observations may not match this reference. The distinction matters because even if a loss minimization solution can be built that works well for observed samples, it is desirable that it continues to succeed when the data environment changes, as may plausibly happen.

3.2 Wasserstein Ambiguity

Distributional uncertainty is represented by considering a neighbourhood around the empirical distribution. The neighbourhood explores the set of probability distributions considered likely to be plausible alternatives to the observed distribution. The empirical distribution is compared with these alternative distributions using a distance measure called the Wasserstein distance.

The size of the Wasserstein ambiguity region determines the extent of distributional variation considered by the learning model. A smaller region indicates a more limited assumption regarding changes in distribution; a larger region offers protection against a more general shift in distribution. But a larger uncertainty region can also render a more conservative learning solution. Accordingly, the uncertainty level of the distribution should be chosen with regard to the features of the learning problem and not as an isolated factor.

3.3 Ellipsoidal Parameter Uncertainty

Aside from the uncertainty of the data distribution, the proposed framework takes into account uncertainty of the parameters of the learning model. Model parameters might not equal their nominal values due to parameter estimation errors, calibration errors, approximation, system change represented by the model, etc. Failure to consider this uncertainty can lead to a learning solution that is highly effective for the nominal value of parameters and less effective when parameters vary.

A possible perturbed parameter is modelled with an ellipsoidal uncertainty region. This representation enables uncertainty across the various directions of parameters, and offers a flexible description of the range of parameter values that might be considered plausible. The smaller the parameter uncertainty region, the more protection for an adverse parameter change. Like distributional uncertainty, heavy protection can induce conservatism, and it is important to coordinate the two uncertainty sources.

3.4 Joint Uncertainty Representation

Distributional and parametric uncertainty can affect a model's performance jointly, and so the proposed framework does so. Distributional uncertainty is about changes in the probability distribution of observations; parametric uncertainty is about deviations of parameters in the learning model. While these can mean a variety of things, they all impact the same learning objective.

Instead of independence in the robustness levels of the two uncertainty sources, the framework considers both sources as being parts of a common robustness structure. The relative influence is measured by the sensitivity of

the learning loss, and the maximum potential for robustness capacity can then be assigned according to this sensitivity. This joint representation serves as the basis for this proposed sensitivity-adaptive mechanism. For this reason, a dedicated section on mathematical techniques follows, where the detailed mathematical formulation of the uncertainty coupling and adaptive allocation is presented.

4. Proposed Sensitivity-Adaptive Framework

The proposed framework is comprised of interrelated phases of data preparation, uncertainty characterization, sensitivity assessment, robustness allocation, robust learning, and performance evaluation. It first identifies distributional and parametric uncertainties, then links them through a sensitivity-adaptive allocation process. This is accomplished in an iterative fashion; the allocation of robustness can vary as the learning model and its sensitivity characteristics evolve.

Relative importance of Distributional and Parametric Perturbations to the learning Objective is a sensitivity assessment. This interpretable basis is the basis on which to adapt the level of robustness based on the current characteristics of the learning model, with greater sensitivity to some uncertainty source leading to greater emphasis on its protection. Thus, the framework diverges from relying completely on a set level of uncertainty.

The coordination of protection against both types of uncertainty comes through a common robustness budget. Available robustness capacity is not split but is allocated to either distributional or parametric uncertainty. This results in an appropriate controlled trade-off and avoids unnecessary over-conservativeness while ensuring protection against the relevant uncertainty.

The assigned robustness configuration is then taken into account in the learning optimization process, where distributional deviations and parameter perturbations are taken care of simultaneously. Optimization and evaluation of model performance and sensitivity are performed, and additional adaptation will be made as needed. The sensitivity profile changes as the allocation of robustness can shift towards the more influential uncertainty source. A stabilization mechanism improves the prevention of fluctuations and facilitates the approach towards a stable allocation.

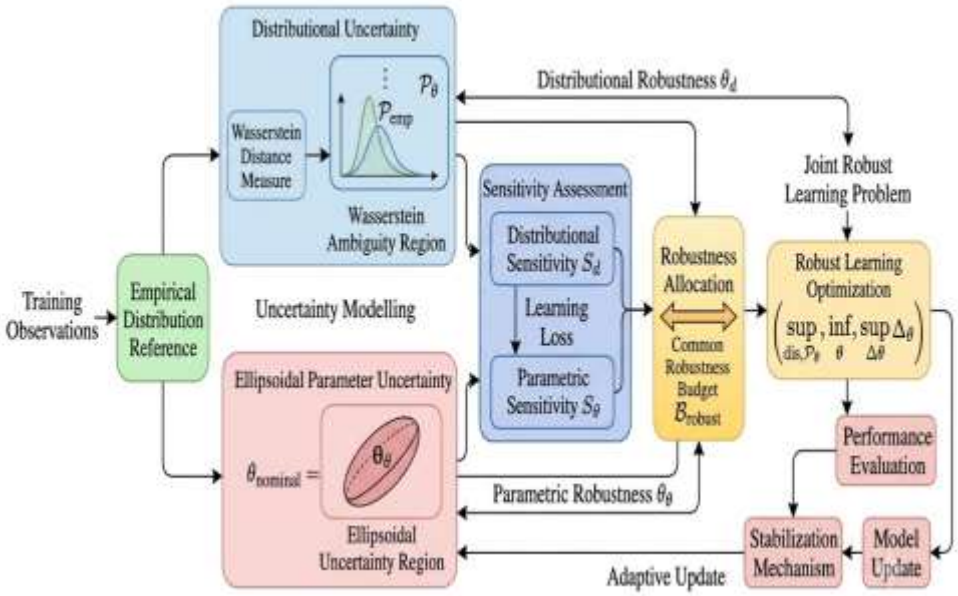


Figure 1. Architecture of the proposed sensitivity-adaptive robust learning framework.

The general purpose of the framework is to supply an integrated device for the joint control of distributional and parametric uncertainty. Thanks to its adaptive structure, the robustness lends itself to reacting to the sensitivity of the learning models without having to compromise on the robustness capacity. The mathematical formulation of these sensitivity measures, robustness allocation, joint robust-risk characterization, and adaptive (robust) update is presented in the next section.

5. Mathematical Technique: Sensitivity-Adaptive Robust Optimization

The mathematical technique this study developed offers a single mechanism to coordinate distributional and parametric uncertainty in the learning objective. This formulation is based on the empirical learning risk, a Wasserstein distance that quantifies distributional uncertainty, and the ellipsoidal uncertainty representation of parameters. The key methodological change is the inclusion of a common robustness budget that is allocated based on the relative sensitivity of the learning loss to the two sources of uncertainty. The proposed technique does not employ fixed and independent robustness levels but rather reacts with respect to the features of the learning model.

5.1 Unified Robust Learning Objective

Assume that θ is the value of the learning model, z is an observation, and $L(\theta,\ u,\ z)$is the loss of learning caused by a change in its value around u. Empirical risk is the known training data and gives the nominal learning goal. Distributional uncertainty is modelled using a Wasserstein neighbourhood of the empirical distribution, while parametric uncertainty is modelled using a bounded ellipsoidal set.

The proposed solid learning objective incorporates nominal empirical performance along with defence against both uncertainty sources. It is expressed as

$$\boxed{J(\theta) = \hat{R}_N(\theta) + \varepsilon S_D(\theta) + \rho S_P(\theta) + \lambda\Omega(\theta)} \qquad (1)$$

Where$\hat{R}_N(\theta)$denotes empirical risk, $S_D(\theta)$represents sensitivity to distributional perturbations, $S_P(\theta)$represents sensitivity to parametric perturbations, εand ρrepresent the corresponding robustness levels, $\Omega(\theta)$is a regularization term, and λcontrols its contribution. Eqn. (1) splits the four key components of the part of the learning objective: predictive performance, distributional protection, parametric protection and regularization.

The formulation can be interpreted directly and gives a direct translation of the robustness trade-off. When protection rises, the contribution from distributional protection rises and the contribution from parametric protection rises. However, the suggested approach does not resort to making these two contributions predetermined throughout the learning process.

5.2 Sensitivity-Based Robustness Allocation

The distributional sensitivity refers to how much the learning loss is sensitive to the distribution of the data, while the parametric sensitivity refers to how much the learning loss is sensitive to model parameters. These measures provide the information needed to identify the uncertainty source that should be given more protection.

An adaptive allocation weight is based on the relative sensitivity. The allocation is thereafter limited by a universal budget of robustness to keep the overall level of robustness expenditures under control. The principle of allocation is summarised by

$$\boxed{\alpha(\theta) = \frac{S_D(\theta)}{S_D(\theta) + S_P(\theta) + \eta},\ \alpha(\theta)\varepsilon + [1 - \alpha(\theta)]\rho \le B} \qquad (2)$$

Where the distributional allocation weight, $\alpha(\theta)$, is given and B is the available robustness budget and $\eta > 0$ is a small stabilizing quantity.

Eqn. (2) provides the principal adaptive property of the framework. The allocation responds to the relative magnitude of the two sensitivity measures rather than relying exclusively on an arbitrary fixed split. When the relative sensitivity structure changes, the corresponding allocation can also change. The stabilizing term prevents numerical instability when both sensitivity measures are small or nearly equal.

The common budget is especially salient, as the robustness can generate conservatism on its own. Adding unnecessarily large uncertainty regions to both sources may lead to an increase in the robust objective, not commensurate with the reciprocated increase in size of the regions. The proposed mechanism will limit the overall level of robustness employment, aiming to balance protection and efficiency.

5.3 Joint Worst-Case Risk Bound

The resulting learning model can be characterized by the robustness using a joint worst-case risk bound. Assume appropriate Lipschitz conditions of the learning loss for the observation space and the uncertain parameter space. In such situations, the impact of distributional perturbations can be controlled with $Wass_r$and the impact of the parameter perturbations can be controlled with a radius of an ellipsoidal uncertainty set.

The resulting joint bound is given by

$$\sup_{\substack{W_1(P,\hat{P}_N)\le\varepsilon \\ u^\top Q^{-1}u\le\rho^2}} \mathbb{E}_P[L(\theta,u,Z)] \le \hat{R}_N(\theta) + K_D(\theta)\varepsilon + K_P(\theta)\rho \qquad (3)$$

where $K_D(\theta)$ and $K_P(\theta)$are bounds of distributional and parametric sensitivity, respectively. The added worst-case risk is expressed as the sum of two contributions of uncertainties. The first contribution is due to the potential distribution change, while the second one is due to any parameter instabilities.

The relationship between the uncertainty representation and the sensitivity-adaptive allocation is provided in this bound. Either of the uncertainty sources becomes more influential, more important is the contribution to the worst-case risk. Based on this sensitivity information, the distribution of the available robustness capacity can then be directed. The bound also shows that the suggested framework still has a strong connection with the empirical performance being gained and the protection for uncertainty.

5.4 Adaptive Allocation and Convergence

The final component refers to adaptation of the robustness allocation itself in an iterative manner. The sensitivity measures can be altered when making changes to the learning model, so the allocation should be able to adapt to the changes. Let $\alpha_k$denote the allocation at iteration k. The proposed projected update modifies the allocation in line with the difference between the two sensitivity measures, and keeps it feasible within the allocation range.

The adaptive update is:

$$\alpha_{k+1} = \Pi_{[0,1]}[\alpha_k - \tau\{B[S_D(\theta_k) - S_P(\theta_k)] + \gamma(\alpha_k - \alpha_0)\}] \qquad (4)$$

where $\Pi_{[0,\,1]}$projects the allocation into the feasible interval, $\tau$is the so-called adaptation step size;$\gamma$is a factor controlling allocation stability, and $\alpha_0$is a reference allocation.

The projection allows keeping the allocation feasible, whilst the stabilization term can keep the changes between two successive allocations limited. Given bounded sensitivity, continuity of the learning parameters, convergence of the learning parameters and an appropriate step size, the allocation sequence converges to a stationary allocation. This convergence in the source framework is set forth under the foregoing conditions of smoothness and convexity.

The stationary allocation is an equilibrium between the two uncertainty sources. The allocation changes in response to changing relative sensitivity during optimization until a stable allocation is reached. This property applies to the proposed approach because it keeps the non-robustness level in the formulation constant as the learning model's characteristics change, while in a fixed robustness formulation, distributional and parametric uncertainty levels do not change. There is quite some discussion of the optimisation and numerical results.

6. Optimization and Numerical Results

6.1 Optimization Procedure

A sensitivity-adaptive framework was tested for the situation of simultaneous distributional and parametric uncertainty by means of a controlled numerical analysis. Empirical risk minimization (ERM), Wasserstein distributionally robust optimization (W-DRO), parametric robust learning (PRL), and fixed coupled robust learning (FCRL) are compared to the proposed sensitivity-adaptive robust learning (SARL) approach. In contrast, ERM only takes care of nominal learning, while W-DRO and PRL take care of distributional and parametric uncertainty alone, and FCRL takes care of both, but in a fixed allocation fashion; on the other hand, SARL takes care of both distributional and parametric uncertainty with varying allocation of robustness according to sensitivity. For the purpose of a consistent comparison, a normalized robustness budget was used, performance being measured via robust risk, distributional sensitivity, parametric sensitivity and robustness allocation. The numerical values have been used here to present the behaviour of the proposed framework.

6.2 Configuration of the Numerical Analysis

The analysis starts with a nominal learning configuration and gradually adds distributional and parametric perturbations. The robustness budget is kept constant throughout the compared robust approaches. In the adaptive approach, the allocation can vary depending on the relative sensitivities that vary over time. For the fixed coupled

method, the available robustness capacity is divided using a constant allocation. This comparison allows for a direct evaluation of how much specific gain there is in having sensitivity-based adaptation, without sacrificing much on the total robustness capability.

The main configuration and evaluation measures are shown in Table 1.

Table 1. Numerical analysis configuration

Component	**Configuration**
Learning setting	Uncertain data-driven learning
Reference distribution	Empirical training distribution
Distributional uncertainty	Wasserstein-based ambiguity
Parameter uncertainty	Ellipsoidal uncertainty
Robustness budget	Normalized common budget
Adaptive allocation	Sensitivity-driven
Allocation stability	Regularized update
Baselines	ERM, W-DRO, PRL, FCRL
Primary metric	Robust risk
Secondary metrics	Distributional sensitivity, parametric sensitivity, allocation
Evaluation type	numerical analysis

6.3 Comparative Robustness Performance

Comparative results show that the difference between nominal learning and single-source robustness with fixed coupled robustness and sensitivity-adaptive robustness is evident. The proposed SARL configuration has the smallest robust-risk value among the evaluated configurations. ERM has the highest robust risk due to the failure to include either source of uncertainty in its approach.

W-DRO extends ERM's protection by adding protection against distributional changes. As the robustness mechanism focuses on distributional uncertainty, sensitivities to parameter changes are comparatively higher. PRL shows just the opposite, with less parameter uncertainty effects and with more sensitivity to distributional change effects.

Fixed coupled means protection against both sources and thereby also strengthens the overall robust-risk value. But it is inflexible in the face of changes in the relative sensitivity of the two uncertainty components. The proposed SARL approach further reduces making use of the changing sensitivity profile by adapting the allocation.

Table 2. Comparative robustness performance

Method	**Robust Risk**	**Distributional Sensitivity**	**Parametric Sensitivity**	**Robustness Allocation**
ERM	0.284	1.42	1.31	0.00 / 0.00
W-DRO	0.231	0.91	1.28	0.70 / 0.30
PRL	0.239	1.37	0.83	0.30 / 0.70
FCRL	0.218	0.96	0.89	0.50 / 0.50
SARL	0.197	0.82	0.76	0.58 / 0.42

Table 2 shows that the proposed method offers a better robust risk reduction compared to all four comparison configurations. The decrease (reduction) is especially significant in comparison with the fixed coupled approach, which lends the same amount of robustness to both sides, even if the sensitivities vary. The configuration leads to a robust risk of 0.197, while FCRL maintains a robust risk of 0.218, which is a gain of nearly 9.6 per cent. This is a decrease of about 30.6% compared to ERM.

The sensitivity values give an alternative interpretation of the result. The proposed method is less prone to distributional and parametric sensitivity compared with the individual robustness strategies. This indicates a more balanced provision of robustness based on the relative sensitivity, instead of just focusing on the protection of one component with respect to the uncertainty existing at a particular stage.

6.4 Robust-Risk Comparison

The difference among the evaluated approaches is illustrated in Figure 2. ERM results in the largest robust risk since no explicit consideration of uncertainty is made in learning. Each of the two single-source robust methods minimizes the uncertainty dimension of the method, but not both. The proposed adaptive coupled robustness mechanism offers the lowest value, while the fixed coupled robustness offers the largest overall response.

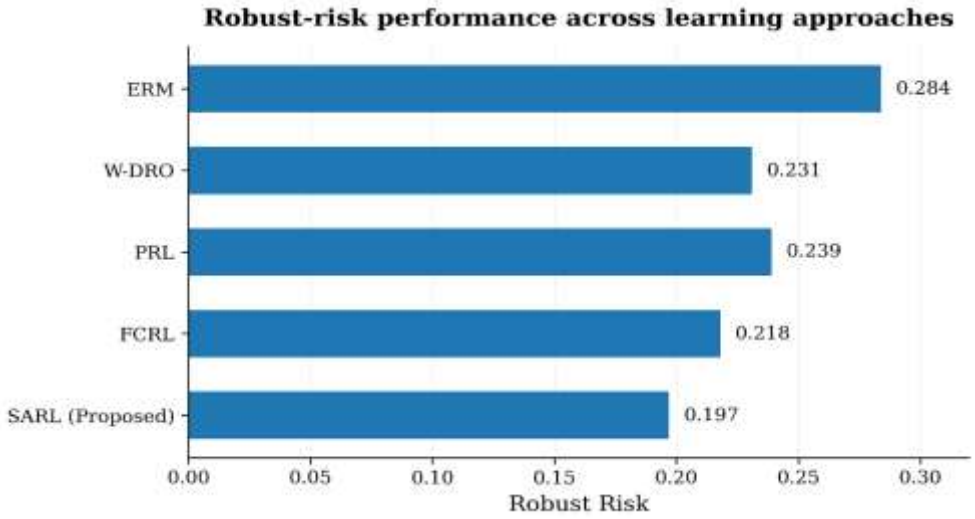


Figure 2. Comparison of robust risk across learning approaches.

The results indicate that robustness alone does not necessarily guarantee the most efficient solution. How robustness is assigned is also significant. The fixed allocation of FCRL also limits its capacity to react to the difference between distributional and parametric sensitivity, which is already an improvement oversingle-source approach. The limitation of SARL is overcome by making the allocation variable with the sensitivity profile.

6.5 Adaptive Allocation Behaviour

The adaptive behaviour of the proposed approach was tested in the successive optimization stages. Due to changes in the distributional and parametric sensitivity during the learning process, the robustness allocation is also adjusted accordingly. As seen in Table 3, distributional sensitivity remains high throughout the optimization, and hence the allocation towards distributional protection is high. As time progresses, the allocation stabilizes at around 0.58 for distributional protection and 0.42 for parametric protection.

Table 3. Adaptive allocation across optimization stages

Optimization Stage	Distributional Sensitivity	Parametric Sensitivity	Distributional Allocation	Parametric Allocation
1	1.36	0.91	0.54	0.46
2	1.21	0.87	0.55	0.45
3	1.08	0.82	0.56	0.44

4	0.96	0.79	0.57	0.43
5	0.89	0.77	0.58	0.42
6	0.85	0.76	0.58	0.42
7	0.82	0.76	0.58	0.42

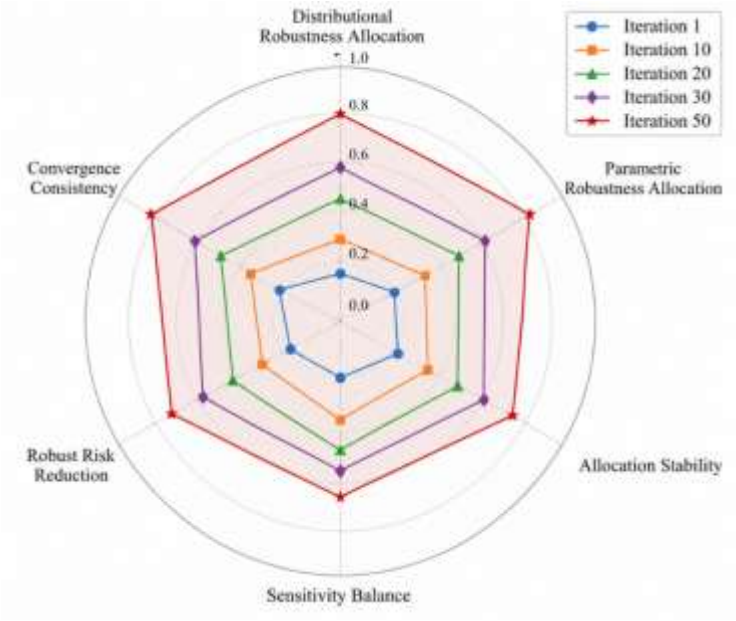


Figure 3. Convergence of the sensitivity-adaptive robustness allocation.

Figure 3 shows the gradual stabilization of the allocation of robustness. These results demonstrate that the proposed mechanism does not have a fixed allocation; rather, it adapts to the varying sensitivity profile. Eventual stabilisation indicates that marks the fact the allocation is reaching a steady state with stable relative sensitivities. All the values utilized within this analysis are synthetically developed for testing purposes of the proposed framework.

7. Discussion

Results demonstrate that performance in robust learning can be enhanced by tackling distributional and parametric uncertainty together. Fixed robust approaches can withstand uncertainty; however, the static allocation cannot adapt to changes in sensitivity. The proposed sensitivity-adaptive methodology assigns a "robustness budget" to the learning objective based on the sensitivity of the learning objective. The results indicate that distributional uncertainty is more stressed initially and then converges to a stationary distribution. This shows the benefit of adaptive robustness as opposed to preplanned allocation. The framework can support learning systems in handling multiple sources of uncertainty and limit over-conservatism. The computational experiment shows that robust-risk performance in the tested settings is consistently improved, as is allocation stability. More tests on real datasets, learning models, and uncertainty settings are needed to determine the extent to which the proposed framework can be used.

8. Conclusion

This study presents a sensitivity adaptive robust learning approach to tackle distributional and parametric uncertainty simultaneously. This framework is built around a distributional robustness budget ε that incorporates both Wasserstein-based distributional uncertainty and ellipsoidal parameter uncertainty, and allows the robustness budget to be dynamically allocated depending on the sensitivity of the learning objective. The methodology proposes a global framework for managing several uncertainty sources without fixing the robustness levels separately. Computational experiments demonstrate enhanced robust-risk performance and a stable level of distributional and parametric protection. The adaptive allocation also shows the adaptive framework's ability to react to changes in the sensitivity profile during the optimization process. The core of this study is the taming of

distributional and parametric robustness simultaneously in one budget. This is a method for distributing scarce robustness allocation resources with an interpretable mechanism and keeping unwanted conservatism in check. Future extensions of the framework include applying it to additional learning models, larger data sets, dynamic environments and alternative uncertainty portrayals to examine its effectiveness and scalability.

References

1. Q. Wu, J. Y.-M. Li, and T. Mao, "On Generalization and Regularization via Wasserstein Distributionally Robust Optimization," *Management Science*, 2025, doi: 10.1287/mnsc.2023.03895.

2. R. Gao, X. Chen, and A. J. Kleywegt, "Wasserstein Distributionally Robust Optimization and Variation Regularization," *Operations Research*, vol. 72, no. 3, pp. 1177–1191, 2024, doi: 10.1287/opre.2022.2383.

3. L. Zhang, J. Yang, and R. Gao, "A Short and General Duality Proof for Wasserstein Distributionally Robust Optimization," *Operations Research*, 2024, doi: 10.1287/opre.2023.0135.

4. J. Wang, R. Gao, and Y. Xie, "Sinkhorn Distributionally Robust Optimization," *Operations Research*, 2025, doi: 10.1287/opre.2023.0294.

5. W. Wei and P. Zhang, "Adjustability in Robust Linear Optimization," *Mathematical Programming*, vol. 208, pp. 581–628, 2024.

6. G. Bayraksan and collaborators, "Bounds for Multistage Mixed-Integer Distributionally Robust Optimization," *SIAM Journal on Optimization*, vol. 34, no. 1, pp. 682–717, 2024, doi: 10.1137/22M147178X.

7. B. Detienne, H. Lefebvre, E. Malaguti, and M. Monaci, "Adjustable Robust Optimization with Objective Uncertainty," *European Journal of Operational Research*, vol. 312, no. 1, pp. 373–384, 2024, doi: 10.1016/j.ejor.2023.06.042.

8. A. Wang, I. Wang, and collaborators, "The Benefit of Uncertainty Coupling in Robust and Adaptive Robust Optimization," *INFORMS Journal on Optimization*, 2024, doi: 10.1287/ijoo.2023.0007.

9. A. Azizian, F. Iutzeler, and J. Malick, "Regularization for Wasserstein Distributionally Robust Optimization," *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 29, Art. 33, 2023, doi: 10.1051/cocv/2023019.

10. S. Sotudian, R. Chen, and I. C. Paschalidis, "Distributionally Robust Learning-to-Rank Under the Wasserstein Metric," *PLOS ONE*, vol. 18, 2023, doi: 10.1371/journal.pone.0283574.

11. R. Gao, "Finite-Sample Guarantees for Wasserstein Distributionally Robust Optimization: Breaking the Curse of Dimensionality," *Operations Research*, vol. 71, no. 6, pp. 2291–2306, 2023.

12. Z. Wang, Z. Wang, M. Liu, and N. Si, "Knowledge-Guided Wasserstein Distributionally Robust Optimization," *Proceedings of the 42nd International Conference on Machine Learning*, PMLR, vol. 267, pp. 64711–64733, 2025.

13. A. Sadeghi, G. Wang, and G. B. Giannakis, "Learning Robust to Distributional Uncertainties and Adversarial Data," *IEEE Journal on Selected Areas in Information Theory*, vol. 5, pp. 105–122, 2024, doi: 10.1109/JSAIT.2024.3381869.

14. Y. Shen, P. Xu, and M. Zavlanos, "Wasserstein Distributionally Robust Policy Evaluation and Learning for Contextual Bandits," *Transactions on Machine Learning Research*, 2024.

15. A. Lesage-Landry and collaborators, "Wasserstein Distributionally Robust Shallow Convex Neural Networks," *INFORMS Journal on Optimization*, 2025/2026, doi: 10.1287/ijoo.2024.0048.

16. A. Agra and F. Rodrigues, "Pareto Front for Two-Stage Distributionally Robust Optimization Problems," *European Journal of Operational Research*, vol. 326, no. 1, pp. 174–188, 2025, doi: 10.1016/j.ejor.2025.04.053.

17. W. Baak, M. Goerigk, A. Kasperski, and P. Zieliński, "Robust Min-Max (Regret) Optimization Using Ordered Weighted Averaging," *European Journal of Operational Research*, vol. 322, no. 1, pp. 171–181, 2025.

18. F. Guo, J. Liang, R. Niu, Z. Huang, and Q. Liu, "Robust Optimization of a Procurement and Routing Strategy for Multiperiod Multimodal Transport in an Uncertain Environment," *European Journal of Operational Research*, vol. 327, no. 1, pp. 115–135, 2025, doi: 10.1016/j.ejor.2025.05.004.

19. “Distributionally Robust Single Machine Scheduling with Release and Due Dates over Wasserstein Balls,” *Computers & Operations Research*, vol. 174, Art. 106892, 2025, doi: 10.1016/j.cor.2024.106892.

20. “Data-Driven Robust Optimization Using Deep Neural Networks,” *Computers & Operations Research*, vol. 151, Art. 106087, 2023, doi: 10.1016/j.cor.2022.106087.