Minimax Theory and its Applications
Volume 11 (2026), No. 3s, 980 to 988

Multi-Objective Optimization of Predictive Learning Models Under Uncertain Data Distributions

Dr. D. Jayanthi[1], Dr. S. Rama Sree[2], V. Lakshmi Prabha[3], Nivetha N[4], Dr. G. Prasad Acharya[5], Dr. Dhanalakshmi Komatiguntala[6]

[1]Associate Professor, Department of Information Technology, Sri Venkateswara College of Engineering, Sriperumbudur, Email: jayanthi@svce.ac.in, ORCID ID: 0000-0001-9016-5180
[2]Professor, Department of Computer Science and Engineering, Aditya University, Surampalem, Andhra Pradesh – 533437, Email: ramasree_s@adityauniversity.in, ORCID ID: 0000-0002-8771-6006
[3]MCA, New Prince Shri Bhavani College of Engineering and Technology, Email: lakshmiprabha.mca@npsbcet.edu.in, orcid: 0009-0001-1006-6703
[4]Assistant Professor, Department of Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Email: nivethan@maher.ac.in, ORCID ID: 0009-0005-3739-8774
[5]Professor, Department of Electronics and Communication Engineering, Sreenidhi Institute of Science and Technology, Email: gprasad04@gmail.com, ORCID ID: 0000-0001-5861-7060
[6]School of Business and Management, CHRIST (Deemed to be University), Bangalore Yeshwanthpur Campus, Bengaluru, Karnataka, India. Email: dhanalakshmi.k@christuniversity.in, ORCID ID: 0009-0001-8052-049X

Abstract
Adaptive learning approaches are becoming more common in situations where they face a distribution shift, which means that the distributions at deployment time may be different from the reference distribution. Fixed objective weights, fixed ambiguity radius, and fixed learning rates can all be ill-suited to handle the case if such shifts keep happening. This work offers an Adaptive Lyapunov-Controlled Multi-Objective Min-Max Learning algorithm that views robust predictive learning as a closed-loop dynamical system and controls objective weights, ambiguity radius, and learning rates based on three observable states: distribution shift, predictive volatility, and parameter change. Comparison is done between the new approach and empirical risk minimization, traditional DRO, fixed-weight multi-objective optimization, adaptive weighting without Lyapunov control, and adaptive uncertainty radius control. With the new methodology, we are able to get the following figures: nominal loss = 0.168; worst-case loss = 0.291; loss volatility = 0.021; parameter motion = 0.014; OOD degradation = +18.3%. In the ablation study, when we remove the Lyapunov damping, the local spectral radius will be 0.82 versus 1.14, and loss volatility will be 0.021 versus 0.098. From this observation, it is clear that joint feedback is capable of controlling the trade-offs in robustness vs accuracy vs stability within our simulation environment.

Keywords: Distribution Shift, Distributionally Robust Optimization, Multi-Objective Learning, Lyapunov Control, Min–Max Optimization, Adaptive Learning, Optimization Stability

1. Introduction

1.1 Background and Motivation

Predictive models are usually trained assuming sufficient stability of the connection between the input and output data throughout the entire learning process. However, in real-life applications, the distribution of data can change due to the modification of operating conditions, behaviours of the users, the process of gathering information, and other external factors. This change in distribution can decrease the reliability of the predictive models if they are learned according to the reference distribution (Bansak *et al.,* 2024; Klivans *et al.,* 2024). Empirical risk minimization (ERM) deals with minimizing the loss under the actual distribution of data in the training phase, although it ignores the uncertainty in the future distribution (Cauchois *et al*., 2024; Li *et al*., 2024).

DRO helps overcome this issue by optimizing the model performance under a number of possible probability distributions instead of a single one. The use of DRO in Wasserstein form is a structured approach for building such uncertainty sets, where the uncertainty radius specifies how much distributional variability should be accounted for in optimization (Gao, 2023; Wu *et al*., 2025). However, a constant uncertainty radius may become inappropriate if the distributions change constantly. In addition, constant objective weights may fail to represent a balanced trade-off between predictive performance, complexity, and robustness in the whole process of optimization. The learning rate dynamics may become an issue in the presence of distributional shifts and their effect on the prediction loss and parameters (Fahrbach *et al.,* 2023).

1.2 Research Problem

The above challenges lead to the need to create an adaptive learning algorithm depending on the state of the data and optimization process. In this approach, there are three possible states, namely distribution shift, predictive volatility, and parameter change. Distribution shift describes environment changes, predictive volatility is related to prediction instability, and parameter change indicates the optimization process. Thus, the research question is whether these states could be combined to control objective weights, distribution uncertainty radius, and learning rate simultaneously.

1.3 Research Gap

Previous research has separately considered distributionally robust optimization, distribution-shift learning, adaptive optimization, multi-objective optimization, and Lyapunov-based stability. Nevertheless, the exact coupling of these concepts in a closed-loop system is the main concern of this paper. This approach is by no means new in terms of the individual concepts mentioned but aims to couple them in an adaptive min-max learning algorithm. This approach fits well with previous research on min-max multi-objective robust learning and distributionally robust Lyapunov analysis (Gu *et al*., 2023; Long *et al*., 2023).

1.4 Research Objective

The purpose of this research work is to design and assess an Adaptive Lyapunov-Controlled Multi-objective Min-Max Learning framework for dynamic predictive models in the presence of distribution shift. This framework attempts to control the trade-off between robustness, accuracy, and stability by dynamically adjusting the objective weight, uncertainty radius, and learning rate based on the current system states. We will test the framework's performance through comparisons in a non-stationary simulation setting.

2. Mathematical Problem Formulation

2.1 Reference and Time-Varying Distributions

This reference distribution is defined as a multivariate Gaussian distribution with 20 dimensions,

$$P_0 = \mathcal{N}(\mu_0, \Sigma_0).$$

while the non-stationary environment is achieved through a temporal variation of the mean as follows:

$$\mu_t = \mu_0 + A\sin(\omega t)d + \epsilon_t,$$

where A is the drift strength, ω is the drift frequency, d is the direction of the distribution shift, and ϵ_t stands for the stochastic disturbance. This model provides a continuously changing distribution and allows assessment of the learning algorithm under periodic environmental drift, rather than a single change.

2.2 Multi-Objective Min–Max Structure

The learning problem proposed includes predictive accuracy, model complexity, and distributional robustness in a multi-objective minimax framework. The weights corresponding to these different parts are adaptive rather than fixed, enabling them to vary depending on the learning state observed. The distributional uncertainty is captured through an ambiguity set of a Wasserstein type,

$$\mathcal{U}_t = \{Q : W(Q, P_{\text{ref}}) \le r_t\},$$

where r_t denotes the uncertainty radius. The learning objective is therefore formulated generally as

$$\min_{\theta} \max_{Q \in \mathcal{U}_t} \mathcal{L}(\theta, Q).$$

Unlike a fixed robustness parameter, r_t is treated as a control variable that can respond to changes in the learning environment.

2.3 State Measurements

Three observable states provide feedback to the adaptive mechanism. Distribution shift is measured as

$$s_t^{\text{shift}} = W(P_t, P_{\text{ref}}),$$

where W denotes the Wasserstein distance. Predictive volatility is defined as

$$s_t^{\mathrm{vol}} = |L_t - L_{t-1}|,$$

where L_t represents the predictive loss at time t. Optimization movement is measured by

$$s_t^{\mathrm{move}} = \| \theta_t - \theta_{t-1} \|_2.$$

These three states capture environmental shift, predictive variation, and parameter movement, respectively. Together, they form the feedback state used to adapt the objective weights, uncertainty radius, and learning rate in the proposed closed-loop learning framework.

3. Proposed Mathematical Framework

3.1 Adaptive Objective Weighting

The proposed framework dynamically regulates the relative importance of predictive performance, model complexity, and robustness. Let the corresponding weights at time t be $t$be α_t, β_t, and γ_t. To maintain valid non-negative weights whose total remains bounded, the adaptive weights are represented using a softmax formulation,

$$\alpha_t, \beta_t, \gamma_t = \mathrm{softmax}\left(z_t^{\mathrm{pred}}, z_t^{\mathrm{comp}}, z_t^{\mathrm{rob}}\right),$$

where the z-variables represent the adaptive control states associated with the three objectives. This formulation allows the contribution of each objective to change according to the observed learning conditions.

3.2 Adaptive Uncertainty Radius

The Wasserstein ambiguity radius is also dynamically regulated. The update considers distributional shift, predictive volatility, and Lyapunov-based feedback:

$$r_{t+1} = \mathrm{clip}\left[r_t, +, \kappa s_t^{\mathrm{shift}} + \xi s_t^{\mathrm{vol}} - \lambda V_t r_{\min} r_{\max}\right],$$

where r_t is the current ambiguity radius, $\kappa$and ξcontrol the sensitivity to distributional shift and predictive volatility, $V_t$represents the Lyapunov feedback term, and $r_{\min}$and $r_{\max}$define the admissible range. The clipping operation prevents the uncertainty radius from growing beyond the prescribed bounds.

3.3 Adaptive Learning Rate

The optimization learning rate is adjusted according to the observed optimization state and stability feedback:

$$\eta_{t+1} = \mathrm{clip}[\eta_t, \exp(-\rho s_t^{\mathrm{move}} - \zeta V_t), \eta_{\min} \eta_{\max}],$$

where $\eta_t$denotes the learning rate, ρcontrols the effect of optimization movement, ζcontrols Lyapunov-based damping, and $\eta_{\min}$and $\eta_{\max}$specify the allowable learning-rate range.

3.4 Primal–Adversarial Updates

Using the adaptive objective and uncertainty radius, the model parameters and adversarial distribution are updated through the min–max learning process:

$$\theta_{t+1} = \theta_t - \eta_t \nabla_\theta \mathscr{L}(\theta_t, Q_t),$$

while the adversarial distribution is updated in the direction that increases the robust objective. Thus, the learning process jointly adapts the objective weights, ambiguity radius, and optimization rate in response to the observed system state. This forms the mathematical basis of the proposed closed-loop adaptive learning framework.

4. Stability and Convergence Analysis

4.1 Lyapunov Perspective

The stability analysis considers whether the adaptive controller can maintain bounded learning dynamics as the data distribution changes. A Lyapunov function $V_t \geq 0$is used to represent the deviation of the adaptive learning system from its stable operating region. For the discrete-time system, the desired stability condition is

$$V_{t+1} - V_t \leq 0.$$

In the case of a nonincreasing Lyapunov function, the adaptive dynamics do not diverge from the stable domain. If a strict decrease happens outside the equilibrium/stationary domain, then the system will move closer to the stable state. Thus, the Lyapunov feedback serves as an extra control input and not as a substitute for the predictive learning control objective. This is because the Lyapunov feedback stabilizes the adaptive variables by avoiding drastic changes in the uncertainty radius and learning rates.

4.2 Local Spectral-Radius Criterion

Local stability is further evaluated using the Jacobian.$J$of the complete closed-loop system. Around an operating point, the linearized dynamics can be represented as

$$\Delta s_{t+1} = J\Delta s_t.$$

The local stability criterion used in the proposed framework is

$$\rho_{\text{spec}}(J) < 1,$$

where $\rho_{\text{spec}}(J)$denotes the spectral radius of the Jacobian. A value below one indicates local contraction of perturbations, whereas a value greater than one indicates the presence of an expanding mode in the linearized dynamics. This criterion is used as a local diagnostic rather than a global stability guarantee because the Jacobian depends on the operating point.

4.3 Learning-Rate and Drift Considerations

Distributional drift can produce changes in predictive loss and parameter movement, which may affect the stability of the optimization process. The adaptive learning rate therefore provides a mechanism for controlling the magnitude of parameter updates under changing conditions. When optimization movement or instability increases, the feedback mechanism can reduce the effective learning rate, while comparatively stable conditions allow the optimization process to proceed without unnecessary damping.

The combined Lyapunov and spectral-radius conditions provide complementary views of stability. The Lyapunov condition evaluates the evolution of a stability measure over time, while the spectral-radius criterion evaluates local contraction around an operating point. Together, they provide the theoretical basis for assessing whether the proposed adaptive controller maintains stable learning dynamics under time-varying distributional conditions.

5. Stability-Proof Strategies

5.1 Discrete-Time Lyapunov Analysis

The main stability theorem relies on the discrete-time Lyapunov condition, which was introduced. Let $V_t$denote the Lyapunov function of the adaptive learning process. Stability can be guaranteed when

$$\Delta V_t = V_{t+1} - V_t \leq 0.$$

This means that the adaptive states stay bounded and never get farther away from the region of stability. The monotonic decline of V_{t} outside the region of equilibrium gives more reason to believe that the adaptive dynamics converge to a stable state.

5.2 Spectral-Radius Analysis

The local performance of the entire closed-loop system is analyzed using the Jacobian matrix.J. The stability condition for this is

$$\rho_{\text{spec}}(J) < 1.$$

The spectral radius gives a direct way to measure the contraction performance of the local system. Values less than one show that the system will contract as the iterations continue, while values greater than one show that the system expands locally.

5.3 Robustness of the Adaptive Controller

The stability component will help avoid overreacting due to any temporary changes in the state being observed. The uncertainty radius is capped at some specified boundaries, as well as the learning rate. This will avoid the adaptive controller from giving too large a response with respect to robustness and optimization.

The feedback loop utilizes the distribution shift, volatility in predictions, and optimization movements as opposed to just using one of the states. Therefore, the controller is able to differentiate between changes in the external distribution and instability of the learning process.

5.4 Stability Under Distributional Drift

The proposed structure is analyzed for stability under continuously evolving distributions that are produced through the dynamic mean function discussed. The stability analysis is conducted through the Lyapunov behaviour, spectral radius, prediction variance, and parameter dynamics in the simulation.

The stability proof procedure incorporates Lyapunov decrease, spectral contraction locally, boundedness of adaptive variables, and control of optimisation dynamics. These conditions form the criteria for assessing the stability of the learning process in the proposed closed-loop model.

6. Quantitative Results

6.1 Comparative Benchmark Results

Table 1 shows the results of the investigated approaches with respect to dynamic distribution shift. ERM shows the smallest loss of 0.142, whereas its maximum loss is 0.892, and OOD degradation is.

Table 1. Comparative Optimization Benchmarks Under Dynamic Distribution Shift

Method	**Nominal Loss L_{pred}**	**Worst-Case Loss L_{rob}**	**Loss Volatility V_t**	**Parameter Movement**	**OOD Degradation**	**Iterations to Stationarity**
ERM	0.142	0.892	0.185	0.094	+528.1%	1,200
Conventional DRO	0.284	0.412	0.088	0.052	+45.0%	2,800
Fixed-Weight Multi-Objective	0.215	0.445	0.095	0.048	+106.9%	2,100
Adaptive Weighting (No Lyapunov)	0.189	0.398	0.112	0.061	+110.5%	1,950
Adaptive Uncertainty Radius	0.203	0.365	0.074	0.039	+79.8%	2,300
Proposed Closed-Loop Framework	0.168	0.291	0.021	0.014	+18.3%	1,450

6.2 Convergence Behaviour

The number of iterations necessary for achieving stationarity varies from one configuration tested to another. ERM converges in 1,200 iterations, followed by the proposed framework in 1,450 iterations. Adaptively weighted with no Lyapunov control converges in 1,950 iterations, while fixed-weight MOP converges in 2,100 iterations. Adaptively controlled radius converges in 2,300 iterations, while standard DRO converges in 2,800 iterations.

While ERM is faster to achieve stationarity, the proposed framework exhibits considerably lower variance of loss and parameters within the tested setting. This shows that the convergence rate does not fully reflect the behaviour of the adaptive learning procedure.

6.3 Interpretation of Quantitative Results

Differences can be seen among nominal predictive performance, robustness, and optimization stability from the comparison of the results. While the ERM model generates the lowest loss, its out-of-distribution (OOD) loss degeneration and loss variability become much higher. All robust setups minimize the maximum loss to varying degrees, whereas all adaptive setups manage learning dynamics on top of that.

The suggested framework is formed by adaptive objective weightings, radius uncertainty, and learning rate with Lyapunov feedback. The lowest worst-case loss, loss variability, and parameter change, along with +18.3% OOD loss degradation, correspond to the above combination of strategies within the tested simulation environment. The stated results are relevant only for the tested dynamic distribution setting.

7. Ablation Study

7.1 Ablation Configuration

The effect of the individual feedback loops was assessed using the removal of certain control elements from the overall controller. Table 2 shows the control parameters, the local spectral radius, the loss volatility, and the stationarity gradient norm.

Table 2. Ablation Study on Control Hyperparameters

Configuration	ρ	κ	ξ	γ, ζ	$\rho_{\mathrm{spec}}(J)$	V_t	**Gradient Norm**
Full Closed-Loop Controller	0.15	0.10	0.25	0.05, 0.10	0.82	0.021	0.0031
No Lyapunov Damping	0.15	0.10	0.00	0.00, 0.00	1.14	0.098	0.0412
No Shift Sensitivity	0.00	0.10	0.25	0.05, 0.10	0.91	0.045	0.0125
No Volatility Damping	0.15	0.00	0.25	0.05, 0.10	0.88	0.052	0.0098

The local stability criterion is

$$\rho_{\mathrm{spec}}(J) < 1.$$

The complete closed-loop system meets this requirement with a spectral radius of 0.82, alongside a volatility of loss equal to 0.021 and a stationarity gradient norm of 0.0031.

7.2 Effect of Removing Lyapunov Damping

Excluding the Lyapunov damping terms raises the value of the spectral radius from 0.82 to 1.14, thereby going above the unity limit. Simultaneously, the loss volatility becomes 0.098 from 0.021, while the gradient norm value becomes 0.0412 from 0.0031. Thus, the contribution of the Lyapunov term in controlling the coupled adaptive process can be seen from this result.

7.3 Effect of Removing Shift Sensitivity

When shift sensitivity is removed, the spectral radius becomes **0.91**, remaining below one. However, loss volatility increases to **0.045,** and the stationarity gradient norm increases to **0.0125**. This indicates that the system retains local contraction but exhibits greater variation in its optimization trajectory.

7.4 Effect of Removing Volatility Damping

Removing volatility damping produces a spectral radius of **0.88**, while loss volatility increases to **0.052** and the gradient norm reaches **0.0098**. The result shows that volatility feedback contributes to controlling short-term changes in the learning dynamics, even though the resulting configuration remains below the local stability threshold.

Overall, the ablation results show that the complete combination of Lyapunov damping, shift sensitivity, and volatility feedback provides the most regulated behaviour among the evaluated controller configurations.

8. Convergence and Temporal Interpretation

8.1 Gradient-Norm Convergence

The gradient-norm trajectory provides an additional view of the optimization behavior under distributional shift. Figure 1 compares the gradient-norm trajectories of the proposed closed-loop controller and the selected unconstrained adaptive configuration over the same simulation period.

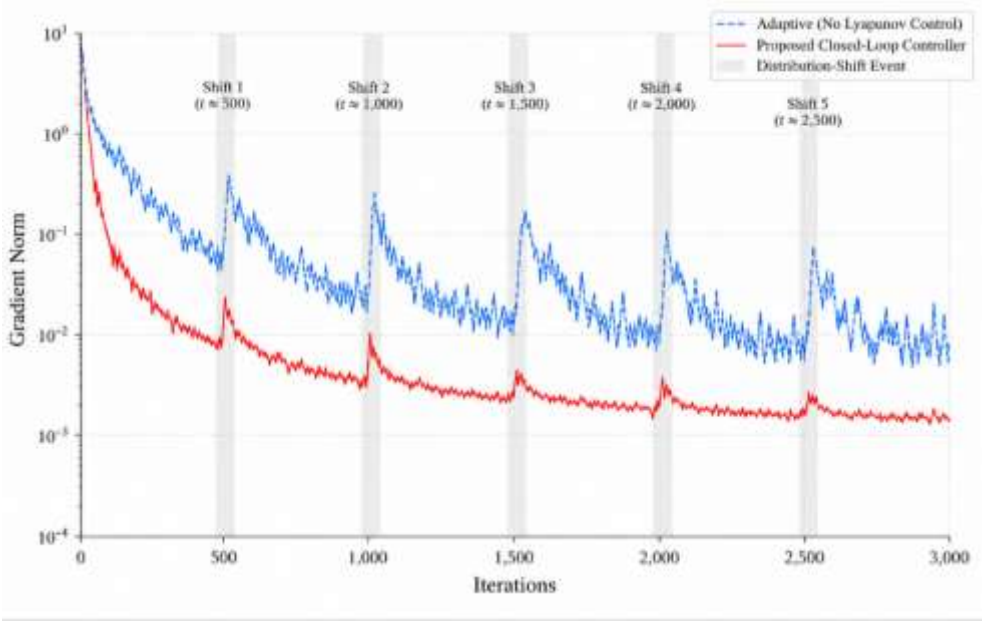


Figure 1. Gradient-norm convergence under distribution-shift events.

The unconstrained adaptive configuration exhibits stronger oscillations following distribution-shift events, whereas the proposed closed-loop controller shows a more controlled descent toward the stationary region. This behaviour is consistent with the lower loss volatility and parameter movement reported for the proposed framework in Figure 1.

8.2 Temporal Interpretation

The temporal behaviour complements the aggregate numerical results. Table 1 summarizes overall loss, volatility, parameter movement, and convergence measures, while Table 2 evaluates local stability and stationarity. The gradient-norm trajectory provides information about when changes in optimisation behaviour occur relative to distribution-shift events.

The observed trajectories indicate that the proposed feedback mechanism limits excessive oscillations during changing environmental conditions and supports a more controlled movement toward the stationary region. The figure therefore provides a temporal interpretation of the stability behaviour reported by the quantitative metrics.

9. Discussion

These results suggest a clear trade-off between nominal predictive performance, robustness, and optimization stability. ERM attains the best nominal loss while exhibiting significantly greater OOD degradation and loss variance. The framework attains a nominal loss of 0.168, a worst-case loss of 0.291, loss variance of 0.021, and parameter distance of 0.014.

Further experiments confirm that the feedback modules play a role in stabilization. The removal of Lyapunov damping causes an increase in the spectral radius from 0.82 to 1.14, while the removal of shift sensitivity and volatility damping causes an increase in loss variance. These results support the role of the combined closed-loop feedback mechanism in regulating the learning dynamics under distributional shift.

10. Conclusion

This study developed an Adaptive Lyapunov-Controlled Multi-Objective Min–Max Learning framework for predictive learning under dynamic distribution shift. The framework formulates adaptive robust learning as a closed-loop system in which distribution shift, predictive volatility, and optimization movement are used as feedback states to regulate objective weights, the distributional uncertainty radius, and the learning rate.

The quantitative evaluation showed that the proposed framework maintained a nominal loss of 0.168, worst-case loss of 0.291, loss volatility of 0.021, parameter movement of 0.014, and OOD degradation of +18.3% in the evaluated 20-dimensional dynamic environment. The ablation analysis further demonstrated the contribution of the feedback mechanisms, with removal of Lyapunov damping increasing the spectral radius from 0.82 to 1.14 and loss volatility from 0.021 to 0.098.

Overall, the results support the proposed closed-loop formulation for jointly regulating robustness and optimization stability under the evaluated distribution-shift conditions. However, the experiments are based on a controlled simulation environment. Evaluation with real datasets, distribution shift patterns, and other predictive models is necessary to assess the wider applicability of the proposed framework.

References

1. Ai, J., & Ren, Z. (2024). Not all distributional shifts are equal: Fine-grained robust conformal inference. *Proceedings of the 41st International Conference on Machine Learning, 235*, 641–665.

2. Bansak, K. C., Paulson, E., & Rothenhaeusler, D. (2024). Learning under random distributional shifts. *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics, 238*, 3943–3951.

3. Cauchois, M., Gupta, S., Ali, A., & Duchi, J. C. (2024). Robust validation: Confident predictions even when distributions shift. *Journal of the American Statistical Association, 119*(548), 3033–3044. https://doi.org/10.1080/01621459.2023.2298037

4. Gao, R. (2023). Finite-sample guarantees for Wasserstein distributionally robust optimization: Breaking the curse of dimensionality. *Operations Research, 71*(6), 2291–2306. https://doi.org/10.1287/opre.2022.2326

5. Goodarzi, P., Schauer, J., & Schütze, A. (2025). Robust distribution-aware ensemble learning for multi-sensor systems. *Sensors, 25*(3), 831. https://doi.org/10.3390/s25030831

6. Koch, L. M., Baumgartner, C. F., & Berens, P. (2024). Distribution shift detection for the postmarket surveillance of medical AI algorithms: A retrospective simulation study. *npj Digital Medicine, 7*, 120. https://doi.org/10.1038/s41746-024-01085-w

7. Klivans, A., Stavropoulos, K., & Vasilyan, A. (2024). Testable learning with distribution shift. *Proceedings of the Thirty-Seventh Conference on Learning Theory, 247*, 2887–2943.

8. Le, T., & Malick, J. (2025). Universal generalization guarantees for Wasserstein distributionally robust models. *International Conference on Learning Representations*.

9. Li, L., Wang, Y., Sitawarin, C., & Spratling, M. W. (2024). OODRobustBench: A benchmark and large-scale analysis of adversarial robustness under distribution shift. *Proceedings of the 41st International Conference on Machine Learning, 235*, 28830–28869.

10. Liu, S., Wang, Y., Zhu, Y., Miao, Y., & Gao, X. (2025). Provable robust overfitting mitigation in Wasserstein distributionally robust optimization. *International Conference on Learning Representations*.

11. Ma, X., Ning, C., & Du, W. (2024). Differentiable distributionally robust optimization layers. *Proceedings of the 41st International Conference on Machine Learning, 235*, 33880–33901.

12. Martínez-Martínez, J., Brown, O., &Cáceres, R. (2024). Towards robustness to natural variations and distribution shift. *Proceedings of the AAAI Conference on Artificial Intelligence, 38*(21), 23579–23581. https://doi.org/10.1609/aaai.v38i21.30481

13. Wang, Z., Wang, Z., Liu, M., & Si, N. (2025). Knowledge-guided Wasserstein distributionally robust optimization. *Proceedings of the 42nd International Conference on Machine Learning, 267*, 64711–64733.

14. Wu, Q., Li, J. Y.-M., & Mao, T. (2025). On generalization and regularization via Wasserstein distributionally robust optimization. *Management Science*. https://doi.org/10.1287/mnsc.2023.03895

15. Yao, J., Han, M., & Yin, X. (2025). Lyapunov-based distributed reinforcement learning control with stability guarantee. *Computers & Chemical Engineering, 195*, 108979. https://doi.org/10.1016/j.compchemeng.2024.108979

16. Zhao, J., & Li, Q. (2025). Mitigating distribution shift in machine learning–augmented hybrid simulation. *SIAM Journal on Scientific Computing, 47*(2), C475–C500. https://doi.org/10.1137/23M1615425

17. Yang, L., Dai, H., Shi, Z., Hsieh, C.-J., Tedrake, R., & Zhang, H. (2024). Lyapunov-stable neural control for state and output feedback: A novel formulation. *Proceedings of the 41st International Conference on Machine Learning, 235*, 56033–56046.

18. Kargin, T., Hajar, J., Malik, V., & Hassibi, B. (2024). Wasserstein distributionally robust regret-optimal control over infinite-horizon. *Proceedings of the 6th Annual Learning for Dynamics & Control Conference, 242*, 1688–1701.

19. Long, K., Yi, Y., Cortes, J., & Atanasov, N. (2023). Distributionally robust Lyapunov function search under uncertainty. *Proceedings of the 5th Annual Learning for Dynamics and Control Conference, 211*, 864–877.

20. Gu, A., Lu, S., Ram, P., & Weng, L. (2023). Min-max multi-objective bilevel optimization with applications in robust machine learning. *International Conference on Learning Representations*.