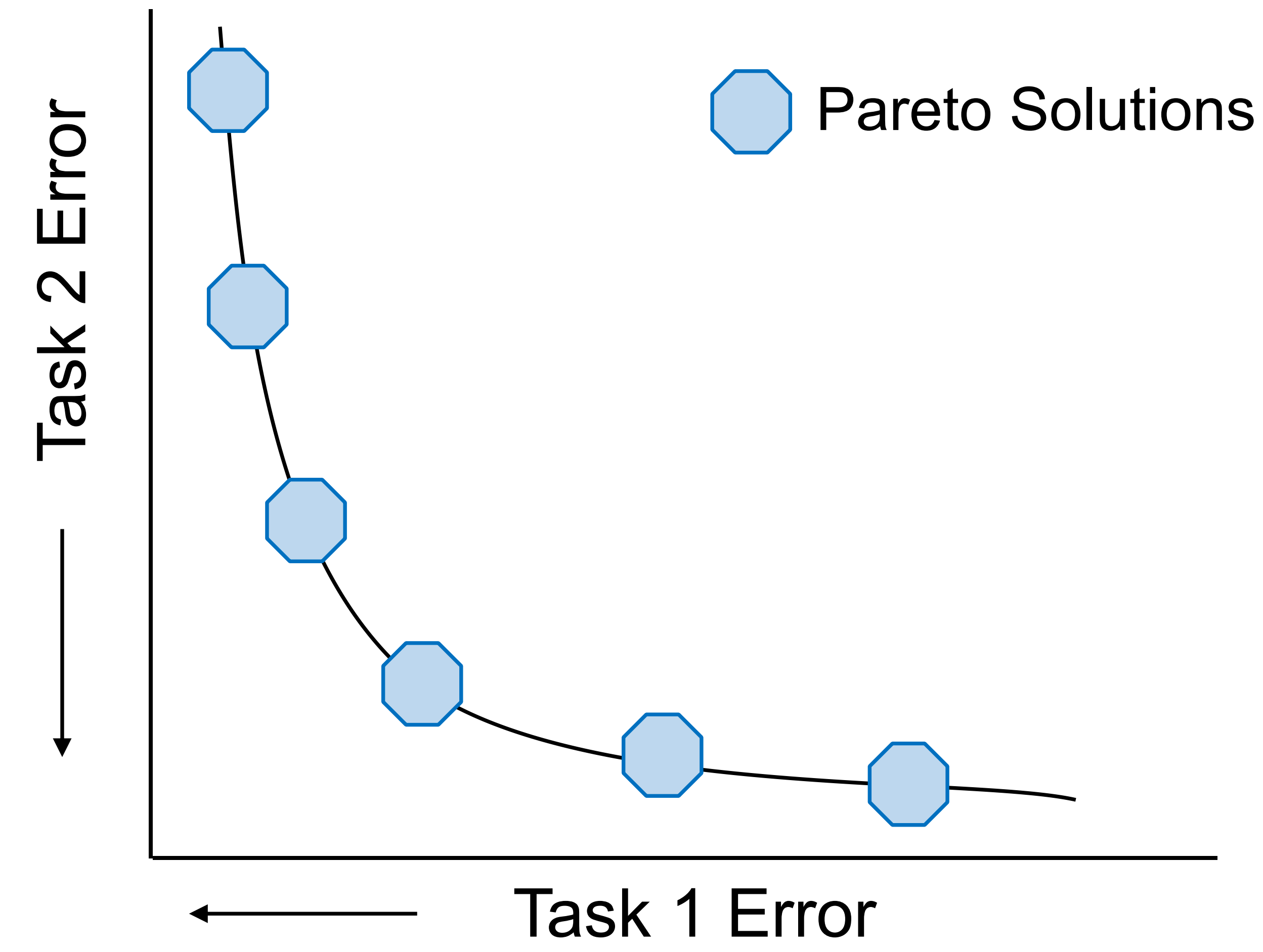


Problem Definition and Contribution

Goal:

To generate a set of widely distributed Pareto solutions with **different optimal trade-offs among tasks** for a multi-task learning (MTL) problem.



Then MTL practitioners can easily select their solutions with preferred trade-offs.

Motivation:

A MTL problem with m tasks is a **multi-objective optimization problem**:

$$\min_{\theta} \mathcal{L}(\theta) = (\mathcal{L}_1(\theta), \mathcal{L}_2(\theta), \dots, \mathcal{L}_m(\theta))^T. \quad (1)$$

Traditional linear scalarization:

$$\min_{\theta} \mathcal{L}(\theta) = \sum_{i=1}^m w_i \mathcal{L}_i(\theta). \quad (2)$$

where w_i is manually chosen by MTL practitioners.

However:

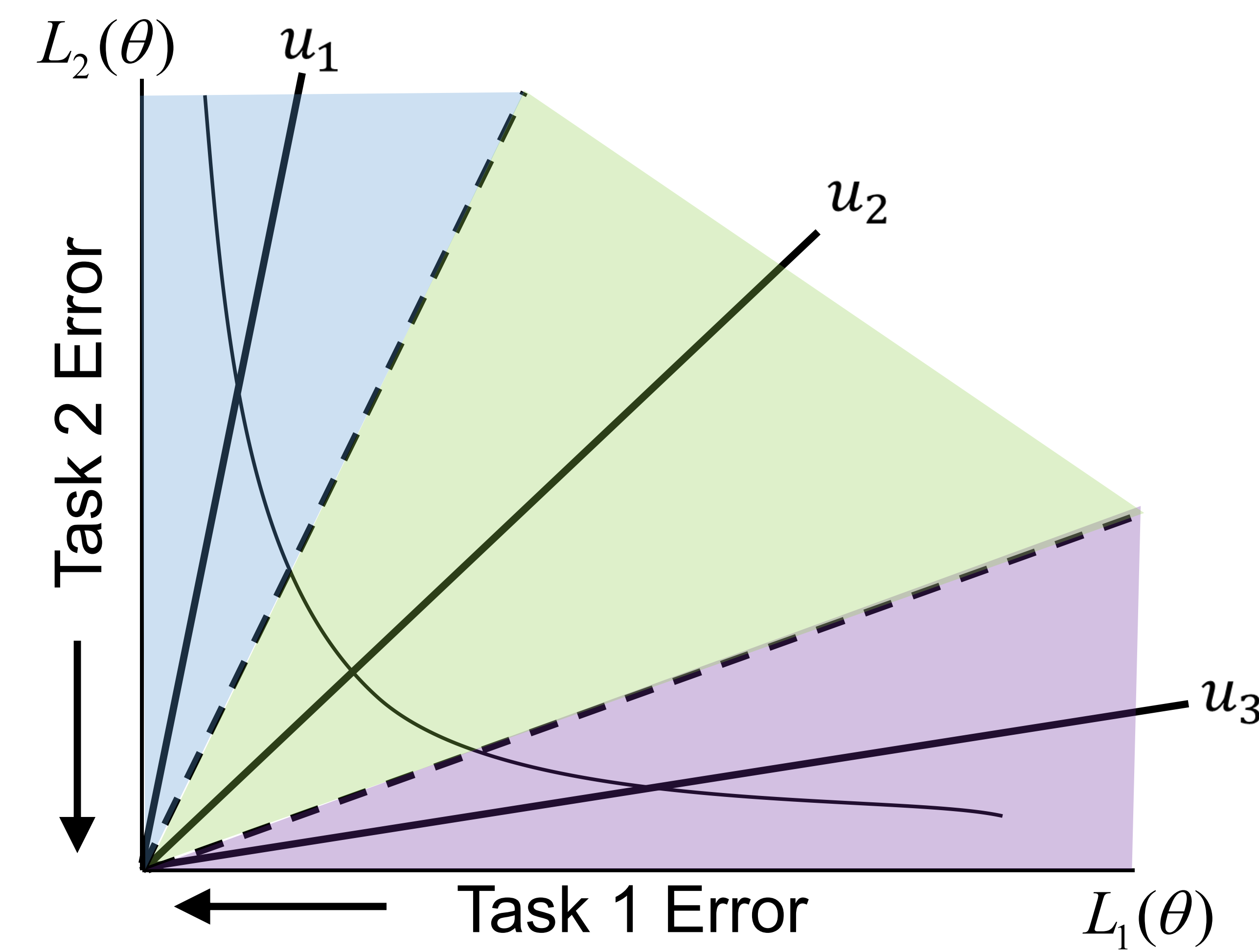
- **Tasks would be conflicted** with each other when the model capacity is limited and fixed, and there is **no single best solution** for all tasks.
- **Exhaustive weights search** is usually needed for finding the solution with particular trade-off that satisfies the practitioner's need.
- Linear scalarization with fixed weight **can not handle concave Pareto front**.
- Current adaptive-weight methods **can not make different trade-offs** among tasks (GradNorm, Uncertainty, MOO-MTL).

Main Contributions:

- **Decompose** a MTL problem into multiple constrained subproblems with different trade-off preferences.
- **Reformulate** each constrained subproblem as a linear scalarization with dynamically adaptive weights.
- **Solve** all constrained subproblems and obtain a set of Pareto solutions with different trade-offs for the MTL.

Pareto Multi-Task Learning

Pareto MTL Decomposition:



Pareto MTL decomposes a MTL problem with a set of unit preference vectors $\{\mathbf{u}_1, \dots, \mathbf{u}_K\}$ into K subproblems:

$$\min_{\theta} \mathcal{L}(\theta) = (\mathcal{L}_1(\theta), \dots, \mathcal{L}_m(\theta))^T, s.t. \mathcal{L}(\theta) \in \Omega_k, \quad (3)$$

where Ω_k is the attracted subregion with smallest acute angle to \mathbf{u}_k :

$$\Omega_k = \{\mathbf{v} \in R_+^m | \mathbf{u}_j^T \mathbf{v} \leq \mathbf{u}_k^T \mathbf{v}, \forall j = 1, \dots, K\} \quad (4)$$

and it can be further simplified as multiple constraints:

$$\mathcal{G}_j(\theta_t) = (\mathbf{u}_j - \mathbf{u}_k)^T \mathcal{L}(\theta_t) \leq 0, \forall j = 1, \dots, K. \quad (5)$$

Solving Each Subproblem: $\theta_{t+1} = \theta_t + \eta d_t$ with a valid direction to minimize all losses and activated constraints:

$$\begin{aligned} (d_t, \alpha_t) = & \arg \min_{d \in R^n, \alpha \in R} \alpha + \frac{1}{2} \|d\|^2 \\ s.t. \quad & \nabla \mathcal{L}_i(\theta_t)^T d \leq \alpha, i = 1, \dots, m. \\ & \nabla \mathcal{G}_j(\theta_t)^T d \leq \alpha, j \in I_{\epsilon}(\theta_t), \end{aligned} \quad (6)$$

where $I_{\epsilon}(\theta) = \{j \in I | \mathcal{G}_j(\theta) \geq -\epsilon\}$ is the index set of activated constraints.

Pareto MTL as Adaptive Linear Scalarization:

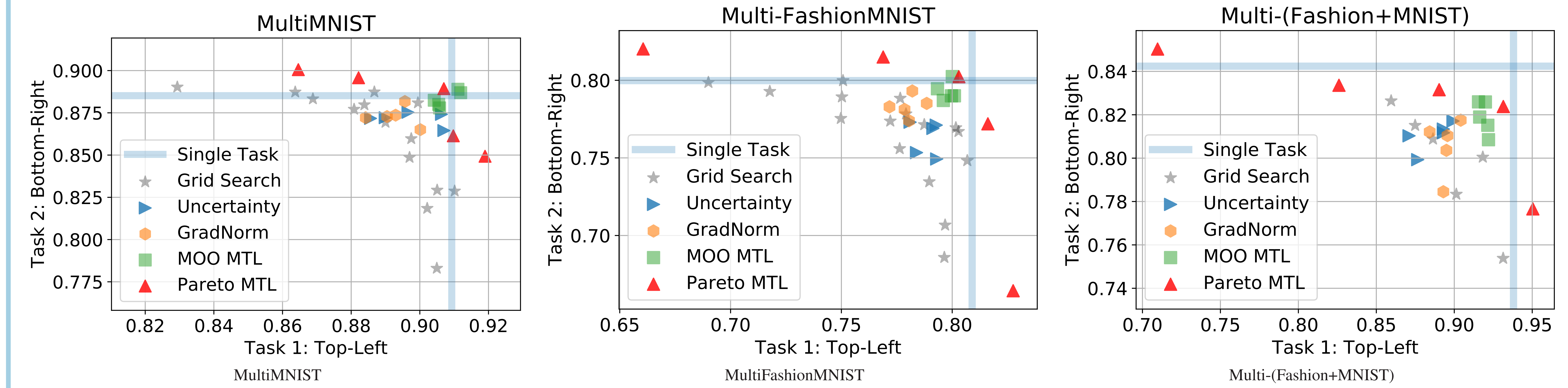
$$\mathcal{L}(\theta_t) = \sum_{i=1}^m \alpha_i \mathcal{L}_i(\theta_t), \quad \text{where } \alpha_i = \lambda_i + \sum_{j \in I_{\epsilon}(\theta)} \beta_j (\mathbf{u}_{ji} - \mathbf{u}_{ki}), \quad (7)$$

where λ_i and β_j are obtained by solving the dual form of (6) with assigned reference vector \mathbf{u}_k at each step.

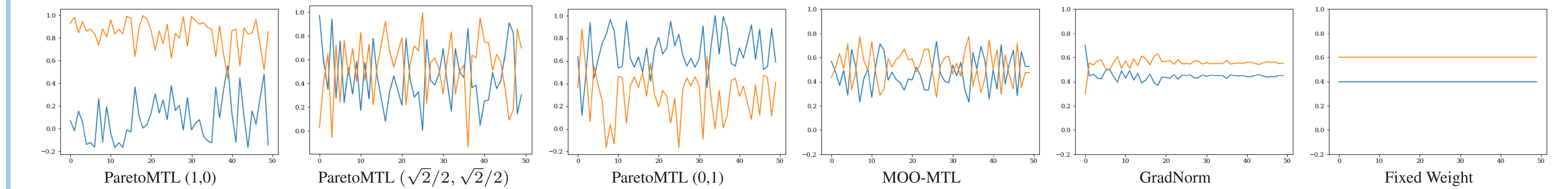
Pareto MTL obtains a set of restricted Pareto optimal solutions with different trade-offs among tasks by solving all subproblems.

Experimental Results

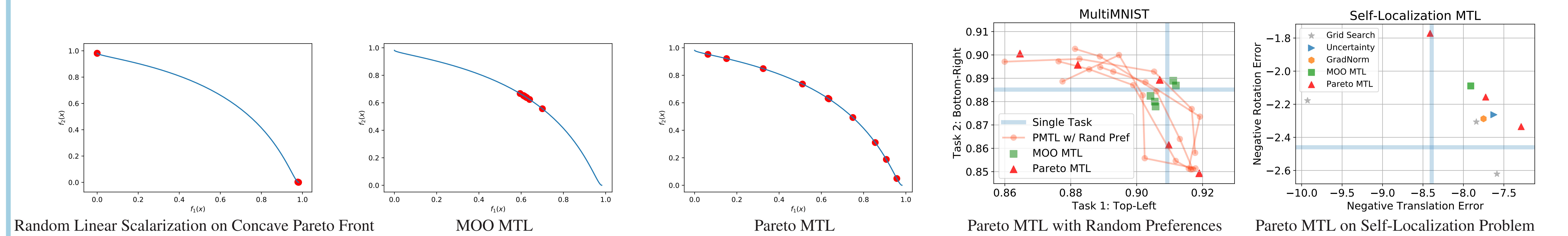
MultiMNIST, MultiFashionMNIST and Multi-(Fashion+MNIST):



Adaptive Weights for Different Algorithms:



More details, discussions, and experimental results can be founded in the paper and supplementary material :



Conclusion

Summary:

We proposed a novel Pareto Multi-Task Learning (Pareto MTL) algorithm to generate a set of well-distributed Pareto solutions with different trade-offs among tasks for a given multi-task learning (MTL) problem.

Key Reference:

Uncertainty: Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. CVPR 2018
GradNorm: Chen et al. Gradnorm: Gradient normal-ization for adaptive loss balancing in deep multitask networks. ICML 2018
MOO MTL: Ozan Sener and Vladlen Koltun. Multi-task learning as multi-objective optimization. NeurIPS 2018

Code & Paper:

