AMARIS: Merging Generalist and Specialist LLMs via Adaptive Subspace Inheritance

Adaptive Model Alignment via Ranked Inherited Subspaces (AMARIS)
Specialist transfer while maintaining preservation budget

Hwee Young Kang · Irwin De Tao Chin
Wen-Haw Chong · Hai Leong Chieu

DSO National Laboratories, Singapore
NeurIPS 2026 - Paris

Motivation: specialization–preservation trade-off

Instruct Following setting, 8B	Core average	IFBench score
Qwen3-8B	64.85	34.33
Light-IF-8B (fine-tuned from Qwen3-8b)	62.32	**44.67**
AMARIS	**67.15**	40.33

Can we keep the core capabilities of base model while inheriting as much specialization as possible?

Core average: AIME26, LCB_v6, IFEval and GPQA:Diamond. Higher is better.

Motivation: specialization–preservation trade-off

Instruct Following setting, 8B	Core average	IFBench score
Qwen3-8B	64.85	34.33
Light-IF-8B (fine-tuned from Qwen3-8b)	62.32	**44.67**
AMARIS	**67.15**	40.33

Can we keep the core capabilities of base model while inheriting as much specialization as possible?

Core average: AIME26, LCB_v6, IFEval and GPQA:Diamond. Higher is better.

Motivation: specialization–preservation trade-off

Instruct Following setting, 8B	Core average	IFBench score
Qwen3-8B	64.85	34.33
Light-IF-8B (fine-tuned from Qwen3-8b)	62.32	**44.67**
AMARIS	**67.15**	40.33

Can we keep the core capabilities of base model while inheriting as much specialization as possible?

Core average: AIME26, LCB_v6, IFEval and GPQA:Diamond. Higher is better.

Motivation: specialization–preservation trade-off

Instruct Following setting, 8B	Core average	IFBench score
Qwen3-8B	64.85	34.33
Light-IF-8B (fine-tuned from Qwen3-8b)	62.32	**44.67**
AMARIS	**67.15**	40.33

Can we keep the core capabilities of base model while inheriting as much specialization as possible?

Core average: AIME26, LCB_v6, IFEval and GPQA:Diamond. Higher is better.

Motivation: specialization–preservation trade-off

Instruct Following setting, 8B	Core average	IFBench score
Qwen3-8B	64.85	34.33
Light-IF-8B (fine-tuned from Qwen3-8b)	62.32	**44.67**
AMARIS	**67.15**	40.33

Can we keep the core capabilities of base model while inheriting as much specialization as possible?

Core average: AIME26, LCB_v6, IFEval and GPQA:Diamond. Higher is better.

Motivation: specialization–preservation trade-off

Instruct Following setting, 8B	Core average	IFBench score
Qwen3-8B	64.85	34.33
Light-IF-8B (fine-tuned from Qwen3-8b)	62.32	**44.67**
AMARIS	**67.15**	40.33

Can we keep the core capabilities of base model while inheriting as much specialization as possible?

Core average: AIME26, LCB_v6, IFEval and GPQA:Diamond. Higher is better.

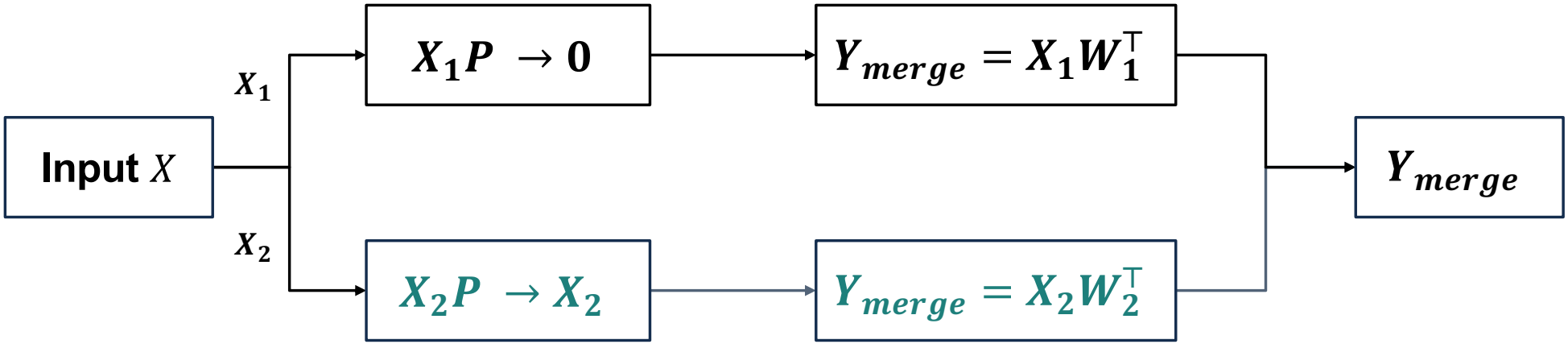
Main Idea: Build an orthogonal projection gate on the specialist transfer

$$\boldsymbol{Y_{merge} = XW_{merge}^{\top} = XW_1^{\top} + XP(W_2 - W_1)^{\top}}$$

W_1 : base weights
X_1: base activations

W_2 : specialist weights
X_2: specialist activations

$P = UU^{\top}$



We want to recover output activations that are good for each model

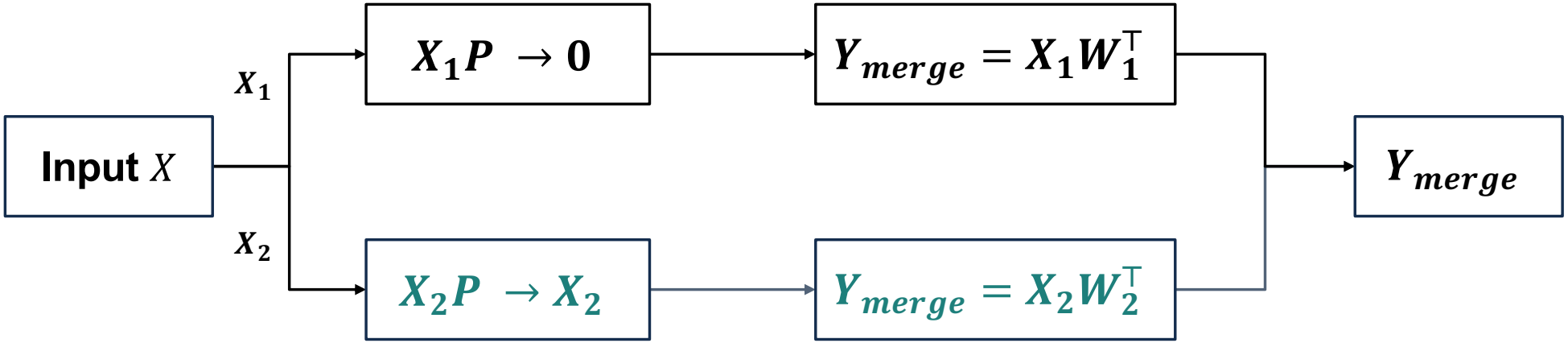
Main Idea: Build an orthogonal projection gate on the specialist transfer

$$\boldsymbol{Y_{merge} = XW_{merge}^{\top} = XW_1^{\top} + XP(W_2 - W_1)^{\top}}$$

W_1 : base weights
X_1: base activations

W_2 : specialist weights
X_2: specialist activations

$P = UU^{\top}$



We want to recover output activations that are good for each model

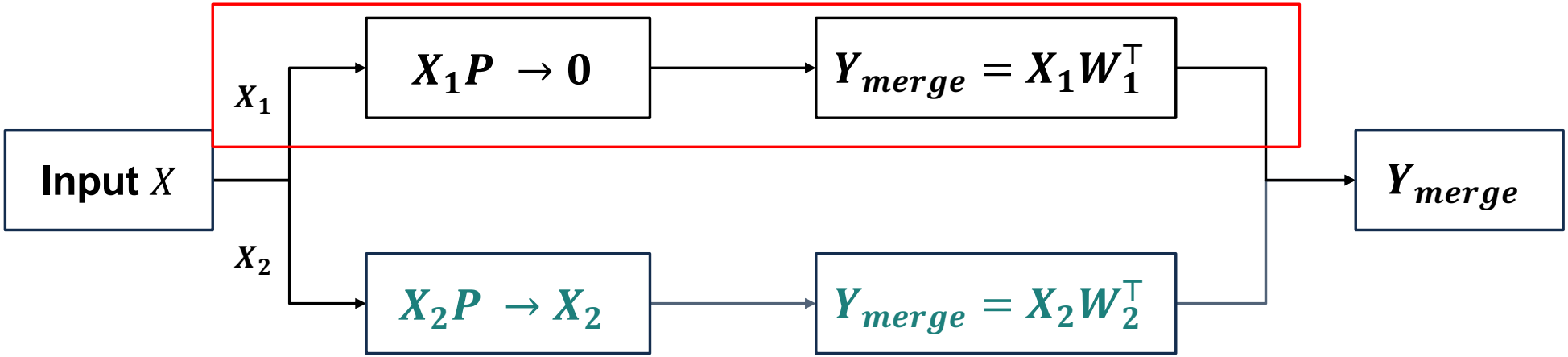
Main Idea: Build a orthogonal projection gate on the specialist transfer

$$\boldsymbol{Y_{merge} = XW_{merge}^{\top} = XW_1^{\top} + XP(W_2 - W_1)^{\top}}$$

W_1 : base weights
X_1: base activations

W_2 : specialist weights
X_2: specialist activations

$P = UU^{\top}$



We want to recover output activations that are good for each model

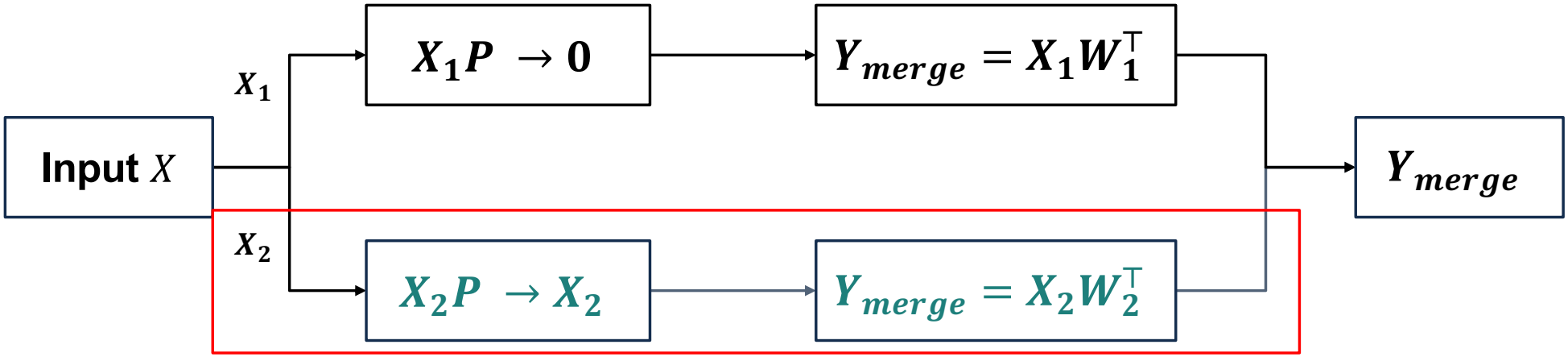
Main Idea: Build a orthogonal projection gate on the specialist transfer

$$\boldsymbol{Y_{merge} = XW_{merge}^{\top} = XW_1^{\top} + XP(W_2 - W_1)^{\top}}$$

W_1 : base weights
X_1: base activations

W_2 : specialist weights
X_2: specialist activations

$P = UU^{\top}$



We want to recover output activations that are good for each model

Objective: Subspace selection under a preservation budget

Estimate activation statistics

Base $C_1 = \frac{1}{N_1} X_1^\top X_1$ and specialist $C_2 = \frac{1}{N_1} X_2^\top X_2$, from small calibration sets

Rank candidate directions

$$\max_{P = P^\top = P^2} tr(PC_2) \quad \text{s.t.} \quad tr(PC_1) \leq \alpha \, tr(C_1)$$

Maximize specialist transfer

Base preservation budget

Smaller α gives a tighter layer-output perturbation budget.

Objective: Subspace selection under a preservation budget

Estimate activation statistics

Base $C_1 = \frac{1}{N_1} X_1^\top X_1$ and specialist $C_2 = \frac{1}{N_1} X_2^\top X_2$, from small calibration sets

Rank candidate directions

$$\max_{P = P^\top = P^2} tr(PC_2) \quad \text{s.t.} \quad tr(PC_1) \leq \alpha \, tr(C_1)$$

Maximize specialist transfer

Base preservation budget

Smaller α gives a tighter layer-output perturbation budget.

Objective: Subspace selection under a preservation budget

Estimate activation statistics

Base $C_1 = \frac{1}{N_1} X_1^\top X_1$ and specialist $C_2 = \frac{1}{N_1} X_2^\top X_2$, from small calibration sets

Rank candidate directions

$$\max_{P = P^\top = P^2} tr(PC_2) \quad \text{s.t.} \quad tr(PC_1) \leq \alpha \, tr(C_1)$$

Maximize specialist transfer

Base preservation budget

Smaller α gives a tighter layer-output perturbation budget.

Objective: Subspace selection under a preservation budget

Estimate activation statistics

Base $C_1 = \frac{1}{N_1} X_1^\top X_1$ and specialist $C_2 = \frac{1}{N_1} X_2^\top X_2$, from small calibration sets

Rank candidate directions

$$\max_{P = P^\top = P^2} tr(PC_2) \quad \text{s.t.} \quad tr(PC_1) \leq \alpha \, tr(C_1)$$

Maximize specialist transfer

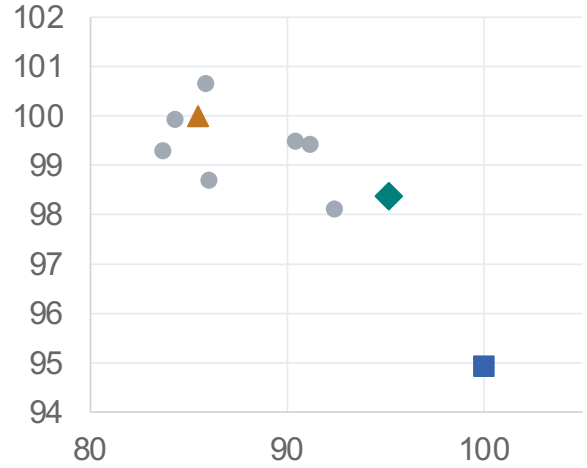
Base preservation budget

Smaller α gives a tighter layer-output perturbation budget.

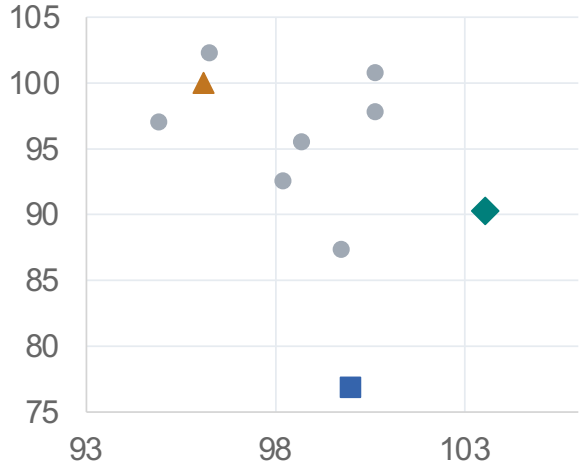
Competitive generalist–specialist trade-offs

Base Specialist Other merges AMARIS

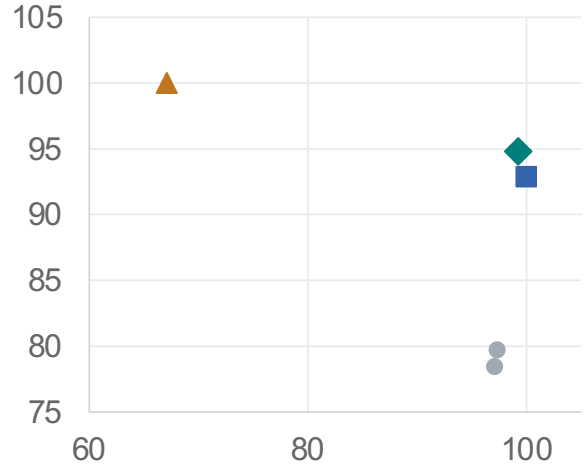
Finance · 8B



Instruction following · 8B



SEA languages · 32B



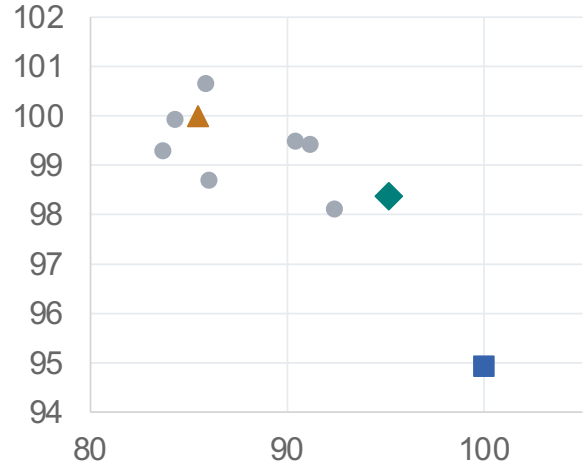
x: Core retention (%) y: Specialist retention (%) Higher and further right is better

SEA: only AMARIS works in improving SEA language performance

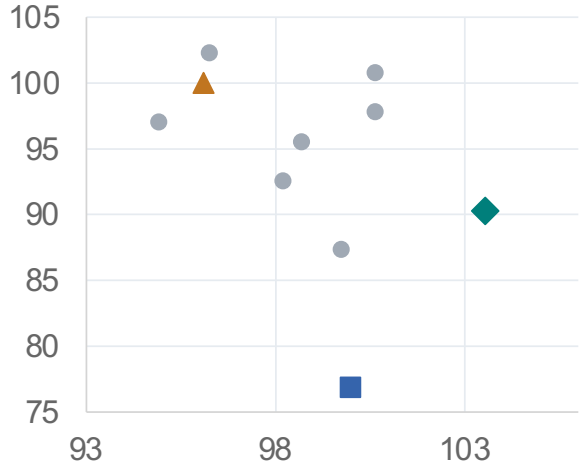
Competitive generalist–specialist trade-offs

Base Specialist Other merges AMARIS

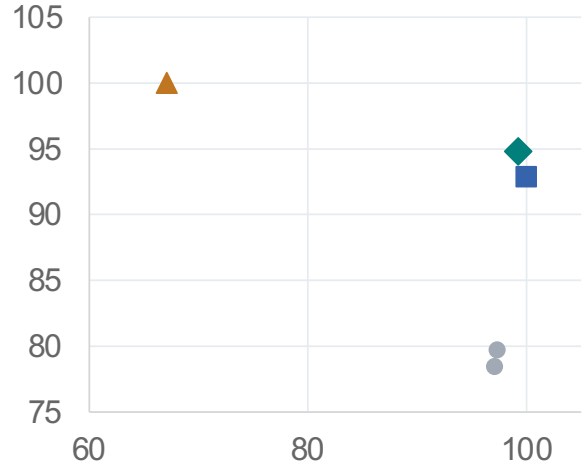
Finance · 8B



Instruction following · 8B



SEA languages · 32B



x: Core retention (%) y: Specialist retention (%) Higher and further right is better

SEA: only AMARIS works in improving SEA language performance

Summary

- No additional training stage
- One static checkpoint for deployment
- One tune-able hyperparameter α

Transfer quality depends on calibration data.

github.com/adriankhy93/AMARIS-neurips