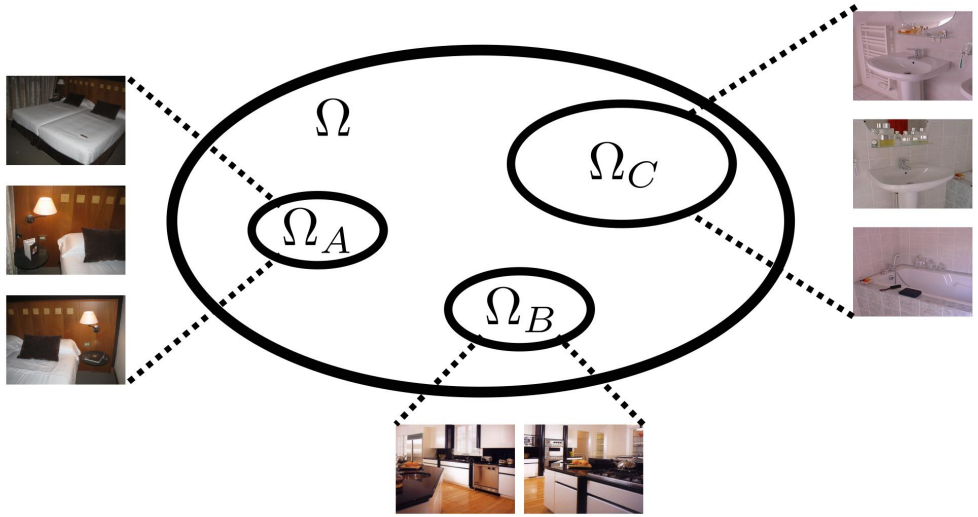
Naturally Constrained Online Expectation Maximization

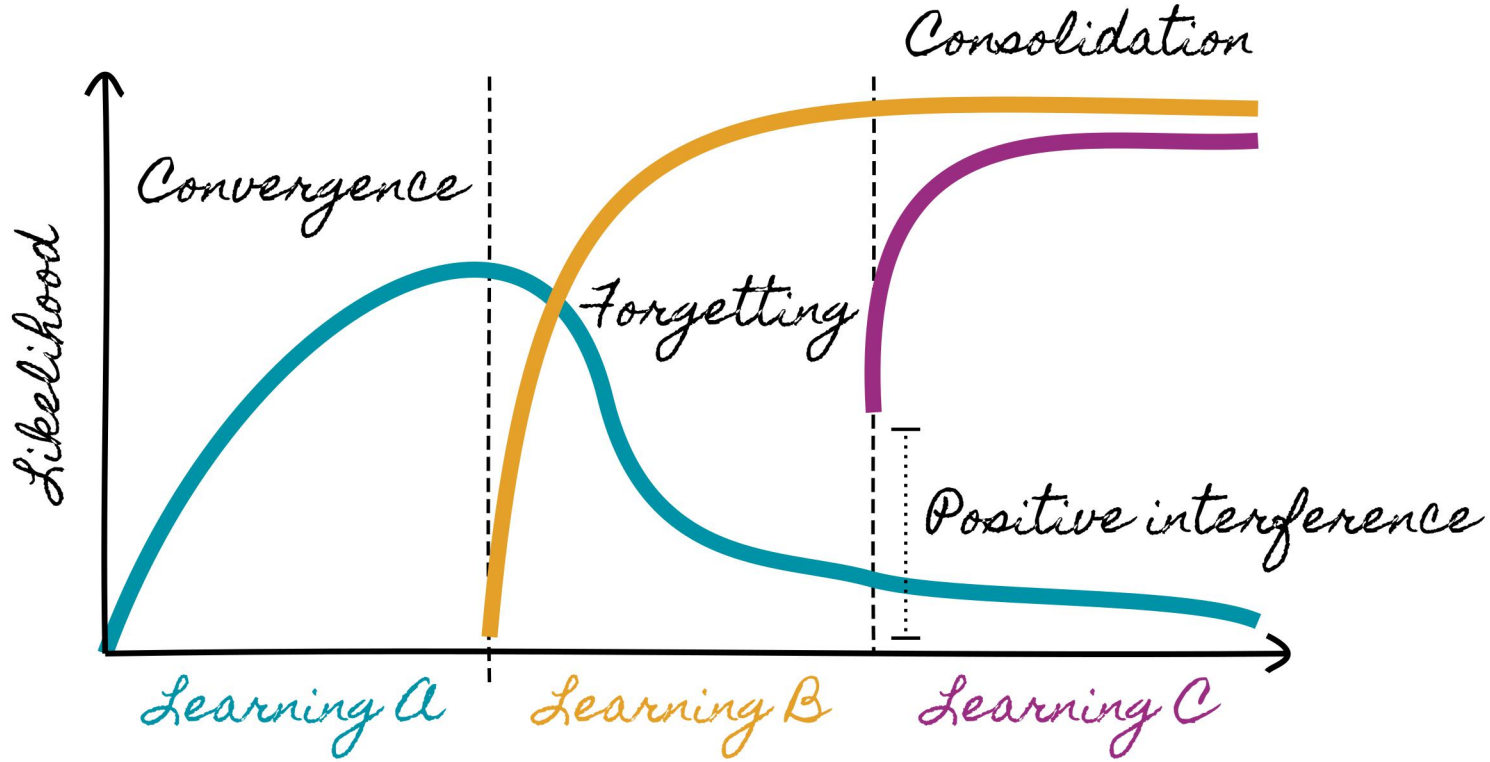
ENSTA
IP PARIS

Motivation: context

Motivation: data



Motivation: Requirements



Methods

Batch	E-step	$Q(\theta, \theta_k) = \mathbb{E}_{Z\|X,\theta_k}\left[\mathcal{L}(X, Z; \theta_k)\right]$
	M-step	$\theta_{k+1} = \arg\max_{\theta} Q(\theta, \theta_k)$
Online	Stochastic	$S^i_{k+1} = S^i_k + \gamma\left(s^i - S^i_k\right)$
Nat-oEM	Regularized	$\theta_{k+1} = \arg\max_{\theta}\left(Q(\theta_k, \theta) - \beta\|\|\theta - \theta^*\|\|_{F^*}\right)$

Evaluation: PPCA

Model:

$$X = WZ + \mu + \varepsilon$$
$$Z \sim \mathcal{N}(0, \mathbb{I})$$
$$\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbb{I})$$

E-step:
Sufficient statistics

$$s^0(x, \mu) = (x - \mu)^T (x - \mu)$$
$$s^1(x, z) = (x - \mu) z^T$$
$$s^2(x, \sigma, M) = \sigma^2 M^{-1} + z z^T$$
$$s^3(x) = x$$

E-step:
Stochastic Integration

$$S^i_{k+1} = S^i_k + \gamma \left(s^i - S^i_k \right)$$

Cappe 2003

Evaluation: PPCA

M-step:

$$\mu_{k+1} = S^3$$

$$W_{k+1} = S^1 {S^2}^{-1}$$

$$\sigma^2_{k+1} = \frac{1}{d}(S^0 - 2\,\mathrm{Tr}(S^1 W^T) + \mathrm{Tr}(S^2 W^T W))$$

R-step:

$$\mu_{k+1\_reg} = \mu_{k+1} - \beta {F^*_\mu}^{-1} (\mu_{k+1} - \mu^*)$$

$$W_{k+1\_reg} = W_{k+1} - \beta {F^*_W}^{-1} (W_{k+1} - W^*)$$

$$\sigma^2_{k+1\_reg} = \sigma^2_{k+1} - \beta {F^*_{\sigma^2}}^{-1} \left(\sigma^2_{k+1} - {\sigma^2}^*\right)$$

Evaluation: PPCA

M-step:

$$\mu_{k+1} = S^3$$

$$W_{k+1} = S^1 {S^2}^{-1}$$

$$\sigma^2_{k+1} = \frac{1}{d}(S^0 - 2\,\mathrm{Tr}(S^1 W^T) + \mathrm{Tr}(S^2 W^T W))$$

R-step:

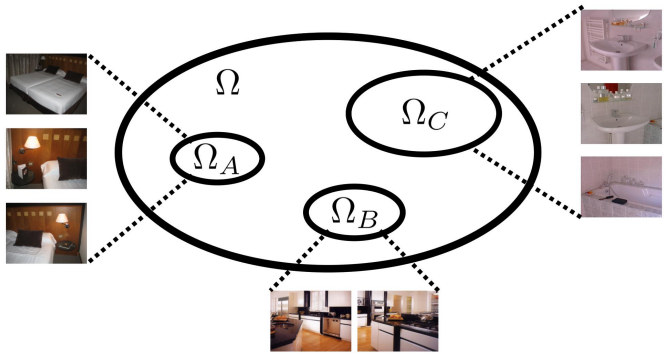
$$\mu_{k+1\_reg} = \mu_{k+1} - \beta {F^*_\mu}^{-1} (\mu_{k+1} - \mu^*)$$

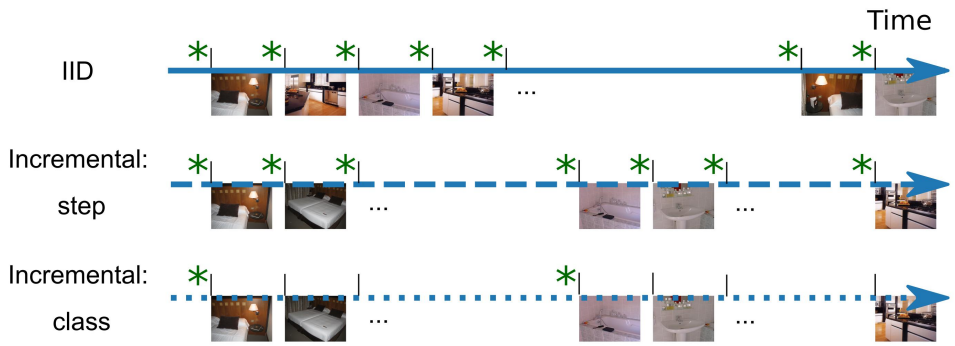
$$W_{k+1\_reg} = W_{k+1} - \beta {F^*_W}^{-1} (W_{k+1} - W^*)$$

$$\sigma^2_{k+1\_reg} = \sigma^2_{k+1} - \beta {F^*_{\sigma^2}}^{-1} \left(\sigma^2_{k+1} - \sigma^{2*}\right)$$

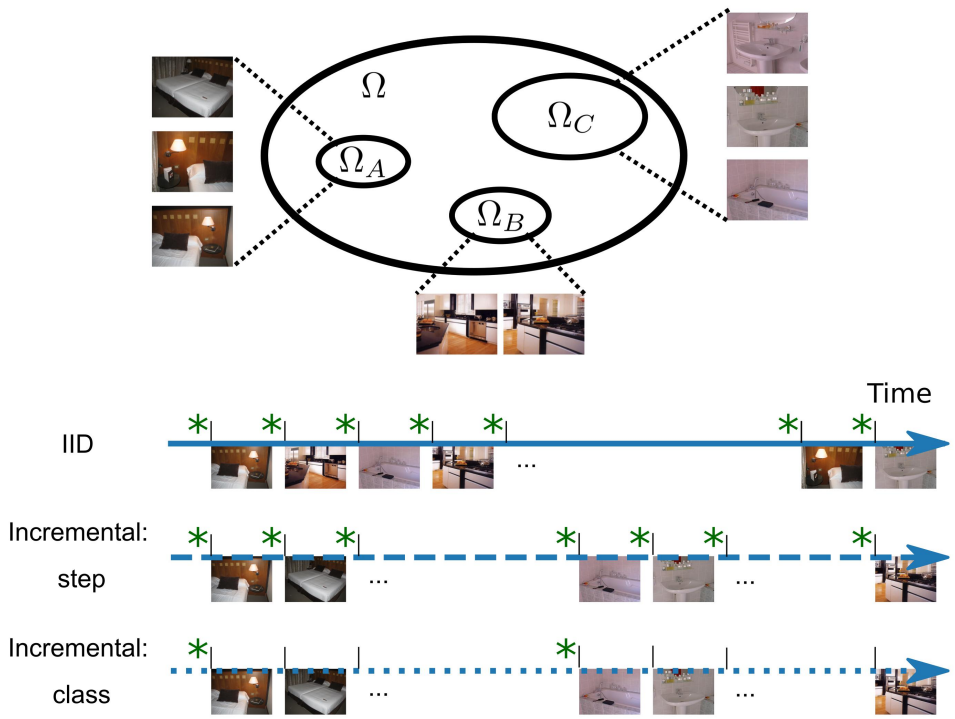
Natural constraint

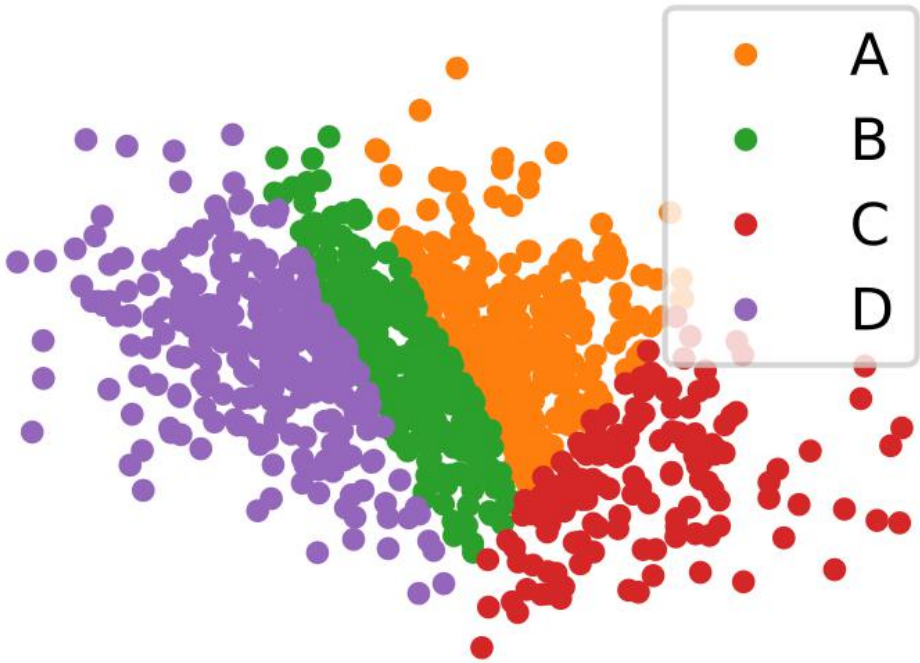
Evaluation: PPCA



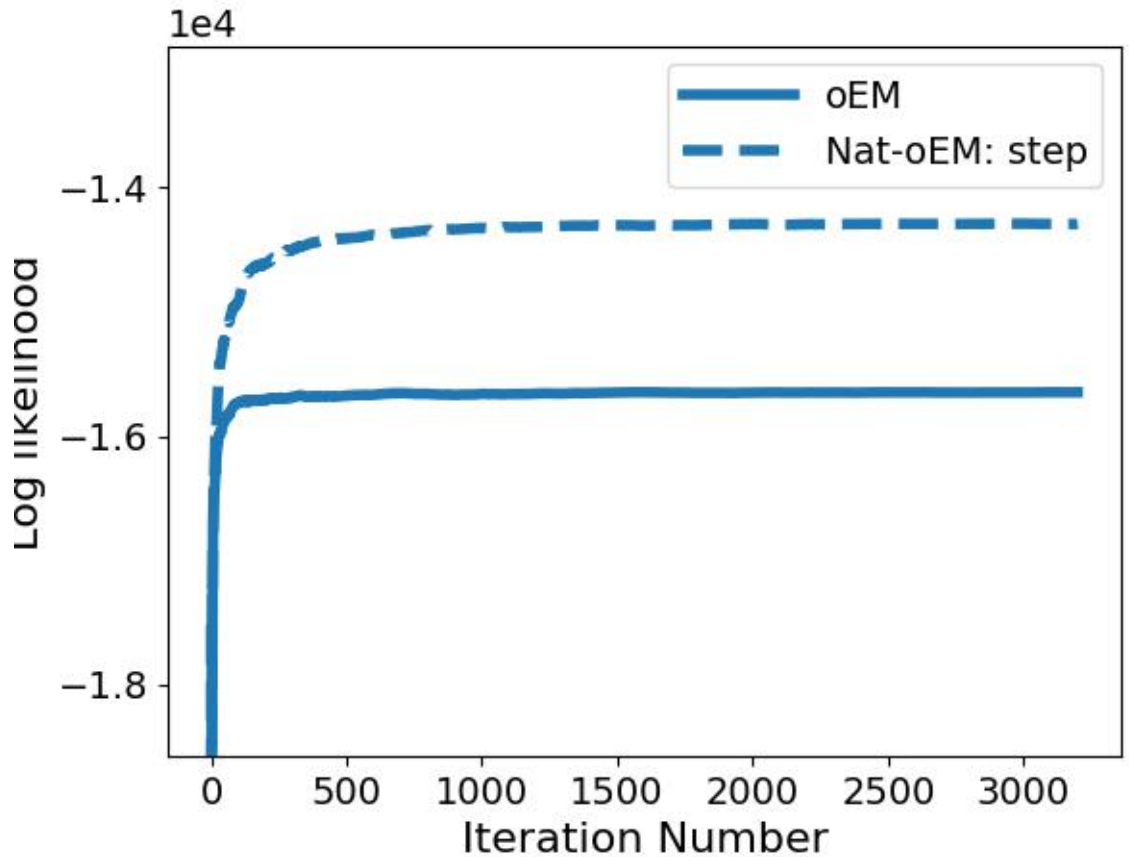


Evaluation: PPCA

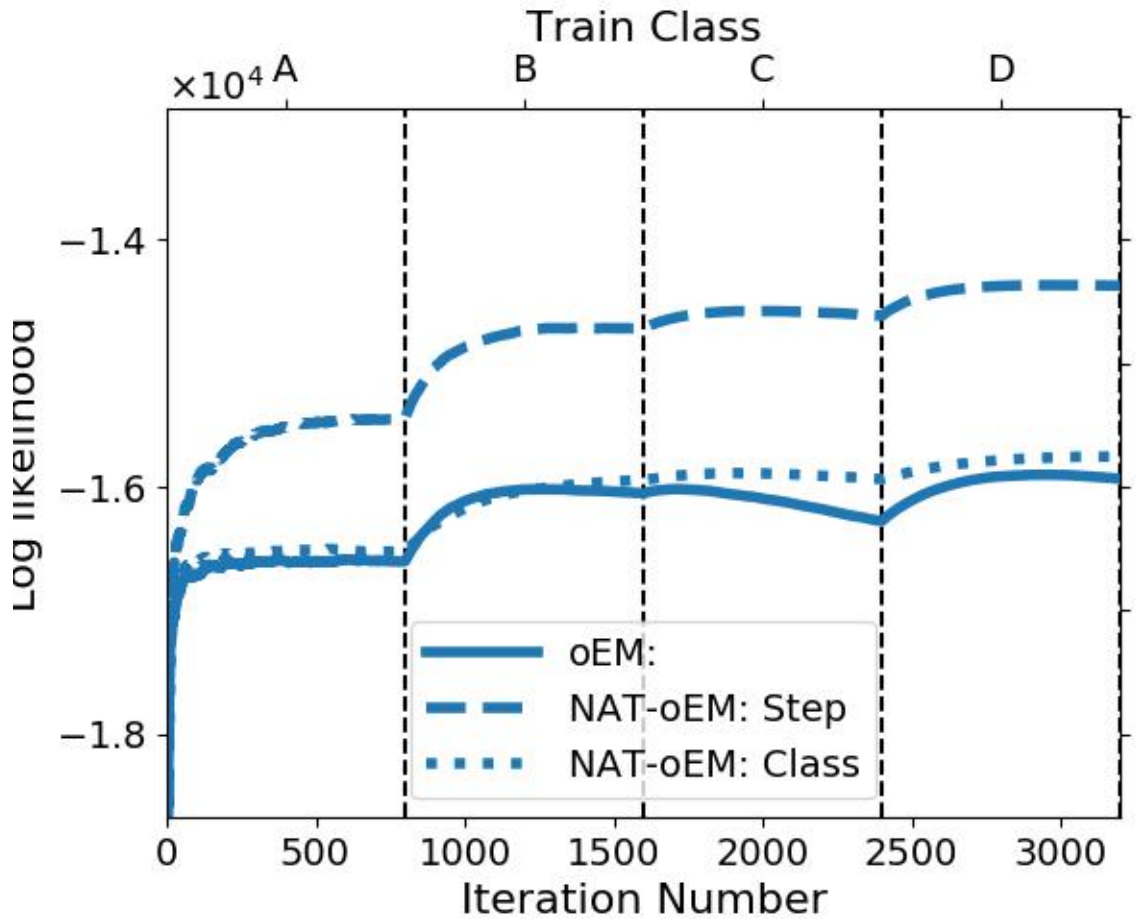




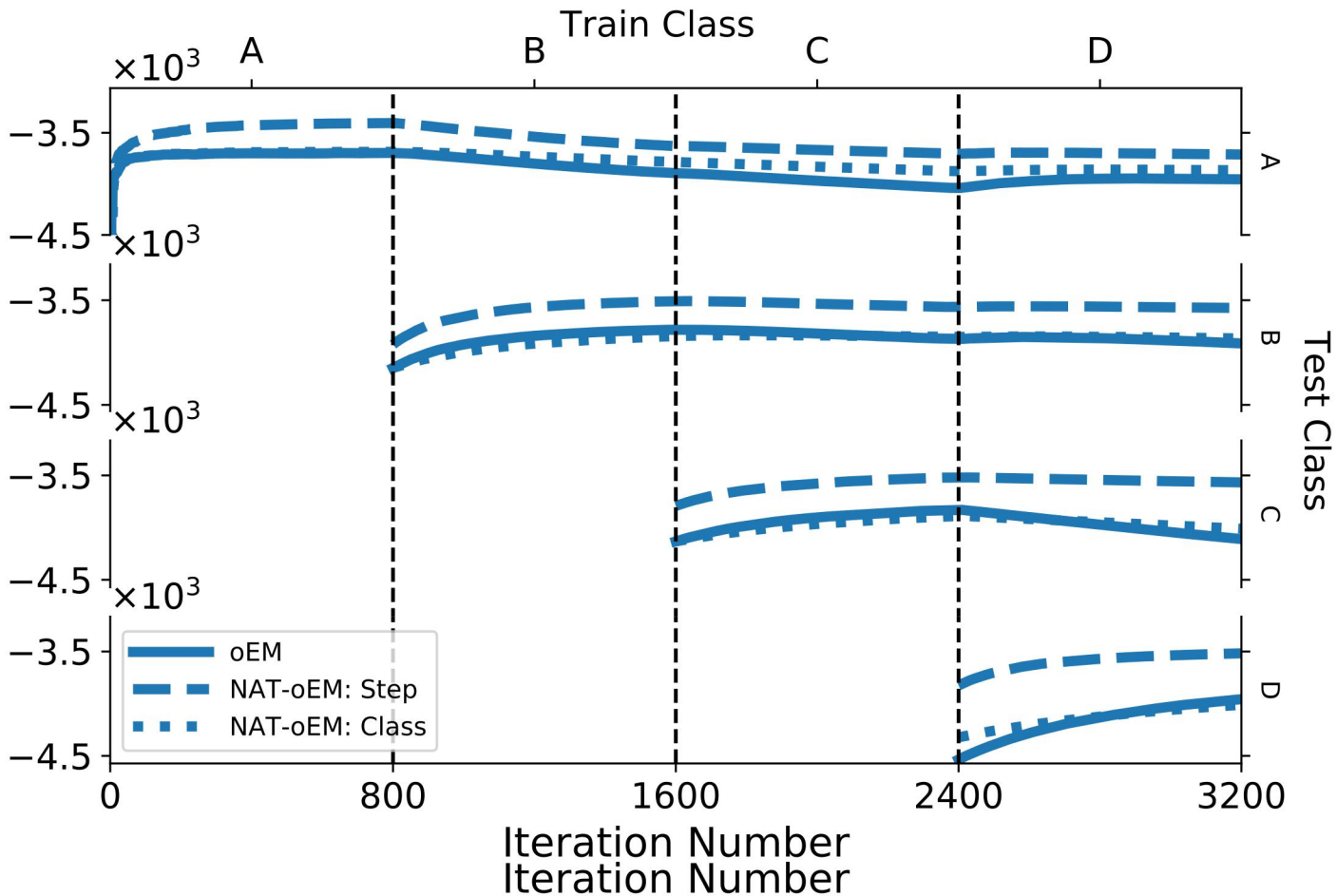
Evaluation: iid - convergence



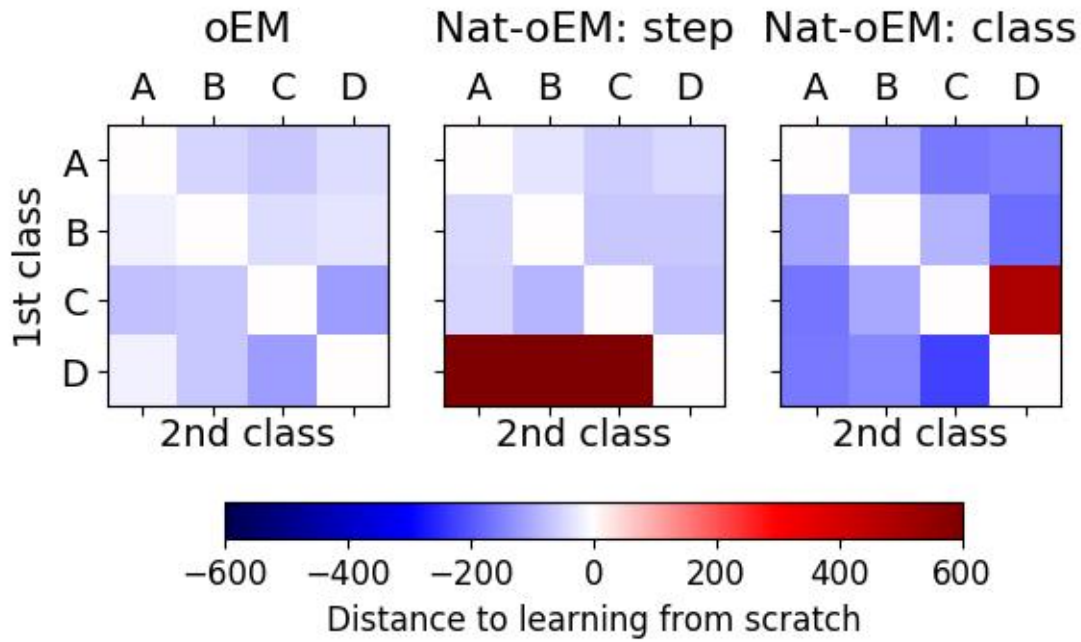
Evaluation: incremental - convergence



Evaluation: incremental - consolidation



Evaluation: incremental - interference



Conclusions:

- Nat-oEM converges faster than oEM,
- Nat-oEM avoids forgetting while oEM does not,
- Overall, Nat-oEM: step performance is better than Nat-oEM class, but increases the computational load,
- Nat-oEM: step introduces positive interference,
- while Nat-oEM: class introduces negative.

Questions?