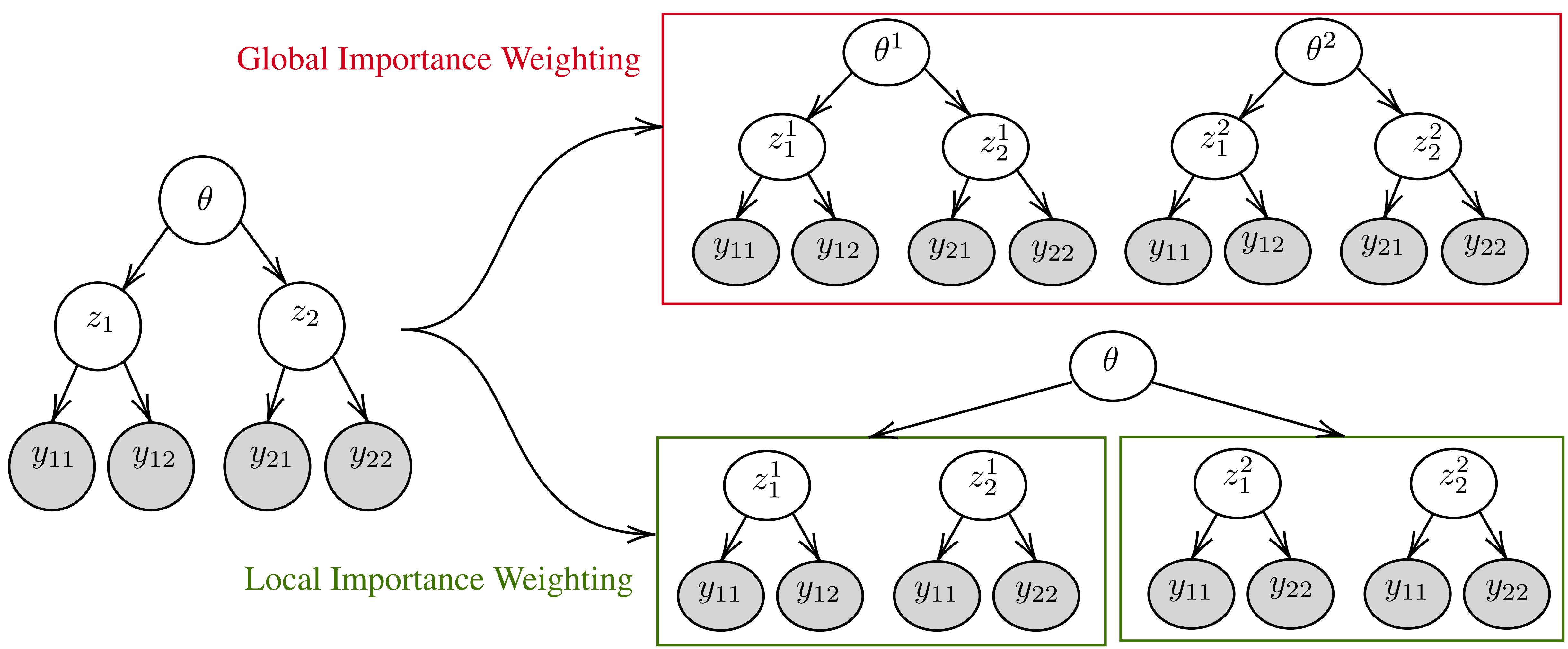
Variational Inference with Locally Enhanced Bounds for Hierarchical Models

Tomas Geffner and Justin Domke, UMass Amherst



Variational Inference

Given $p(x, z)$ find $q(z) \approx p(z|x)$

Variational Inference

Given $p(x, z)$ find $q(z) \approx p(z|x)$

Maximize $\text{ELBO} = \mathbb{E}_{z \sim q(z)} \log \frac{p(z, x)}{q(z)} \leq \log p(x)$

Variational Inference

Given $p(x, z)$ find $q(z) \approx p(z|x)$

Maximize $\mathrm{ELBO} = \mathbb{E}_{z \sim q(z)} \log \frac{p(z, x)}{q(z)} \leq \log p(x)$

Equivalent to minimizing $\mathrm{KL}(q(z) \| p(z|x))$

Monte Carlo VI

Draw multiple samples, get a better bound (Burda et al. 2016).

$$\mathrm{IW-ELBO} = \mathbb{E} \log \frac{1}{K} \sum_{k=1}^{K} \frac{p(z^k, x)}{q(z^k)}$$

Monte Carlo VI

Draw multiple samples, get a better bound (Burda et al. 2016).

$$\mathrm{IW-ELBO} = \mathbb{E} \log \frac{1}{K} \sum_{k=1}^{K} \frac{p(z^k, x)}{q(z^k)}$$

Or, can use Annealed Importance Sampling (AIS) to generate $z^1, \cdots, z^K$ and then define:

$$\mathrm{AIS-ELBO}$$

(Wu et al. 2020, Thin et al. 2021, Zhang et al. 2021, Geffner and Domke 2021)

Monte Carlo VI

Advantages:

- Bound gets tighter as number as K increases.

Monte Carlo VI

Advantages:

- Bound gets tighter as number as K increases.

Problems with big models:

- Monte Carlo struggles in high dimensions (less tightening).
- Incompatible with subsampling (expensive).

Monte Carlo VI

Advantages:

- Bound gets tighter as number as K increases.

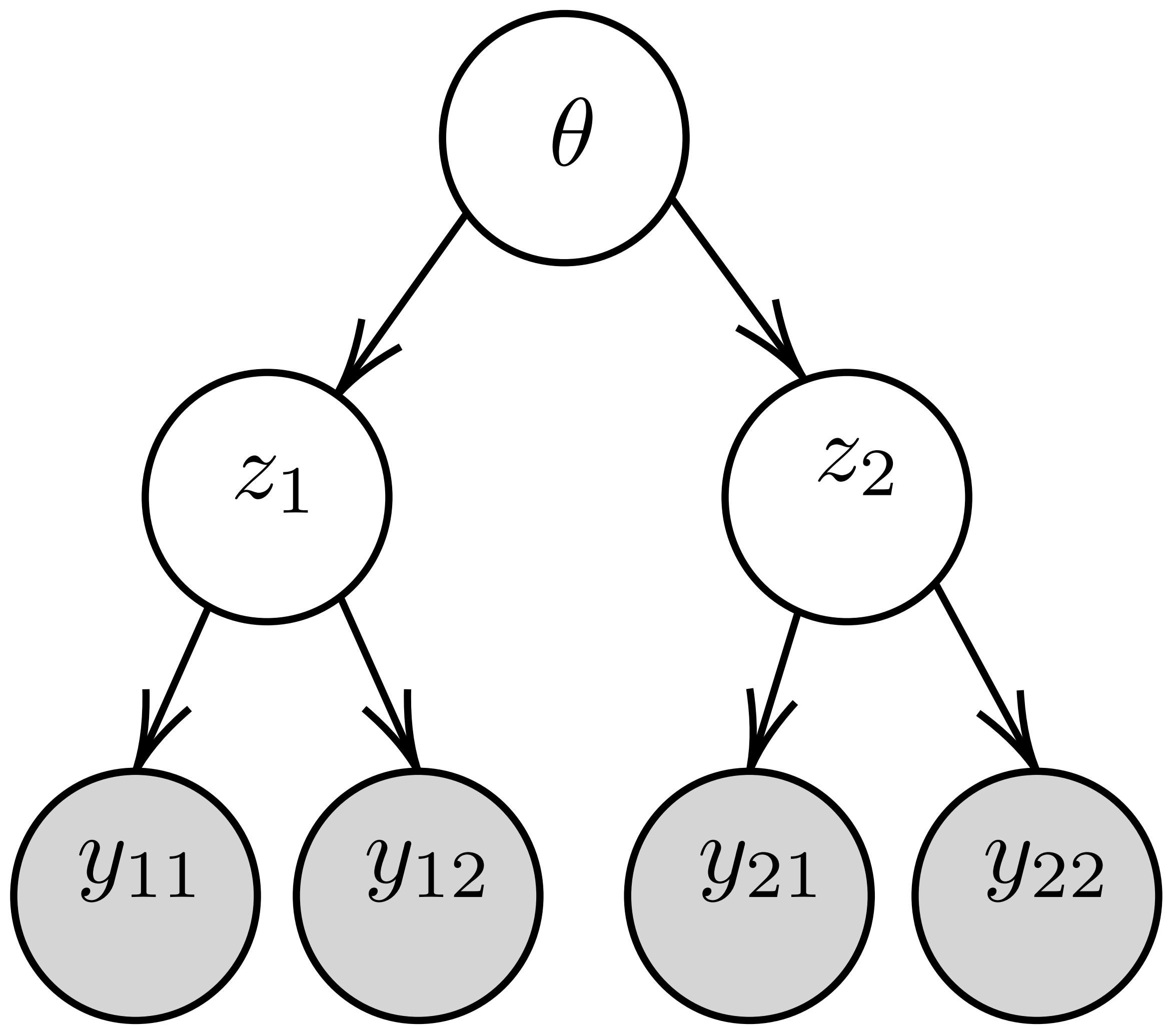
Problems with big models:

- Monte Carlo struggles in high dimensions (less tightening).
- Incompatible with subsampling (expensive).

This paper: Can we solve these problems?

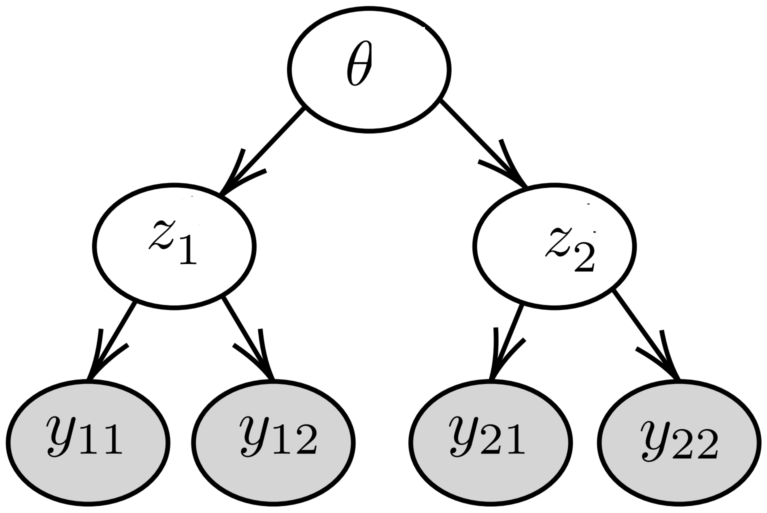
Heirarchical Models

$$p(\theta, z, y) = p(\theta) \prod_{i=1}^{M} p(z_i|\theta)p(y_i|z_i, \theta)$$



Regular VI

$$\text{ELBO} = \mathbb{E} \log \frac{p(\theta, z, x)}{q(\theta, z)}$$

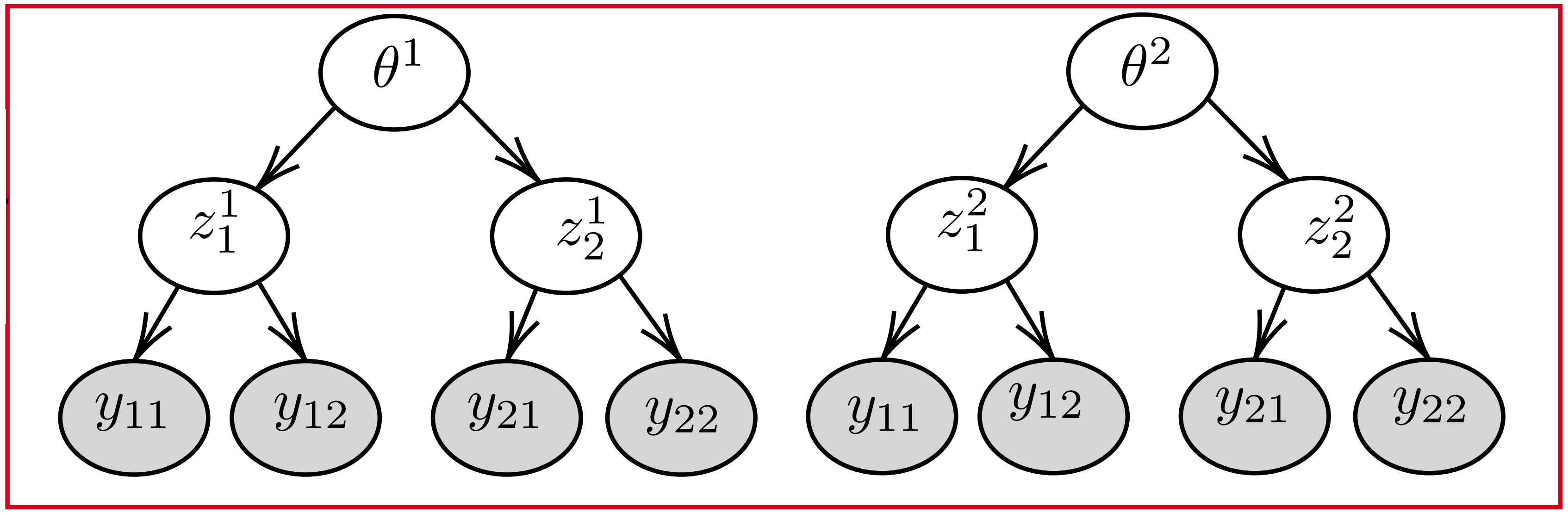


Subsampling: ✓

Enhanced bound: ✗

Global IW

$$\mathrm{IW-ELBO} = \mathbb{E} \log \frac{1}{K} \sum_{k=1}^{K} \frac{p(\theta^k, z^k, x)}{q(\theta^k, z^k)}$$

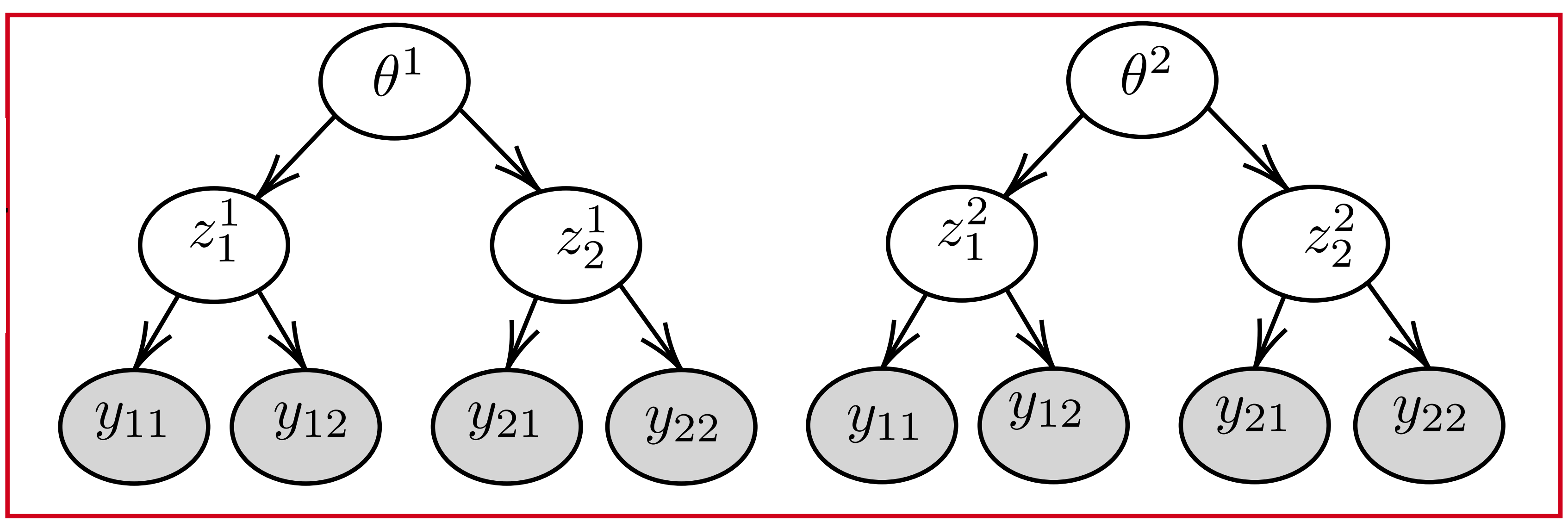


Subsampling: ✗

Enhanced bound: ✓

Global IW

$$\mathrm{IW-ELBO} = \mathbb{E}\log\frac{1}{K}\sum_{k=1}^{K}\frac{p(\theta^k)}{q(\theta^k)}\prod_{i=1}^{M}\frac{p(z_i^k, x_i|\theta^k)}{q(z_i^k|\theta^k)}$$



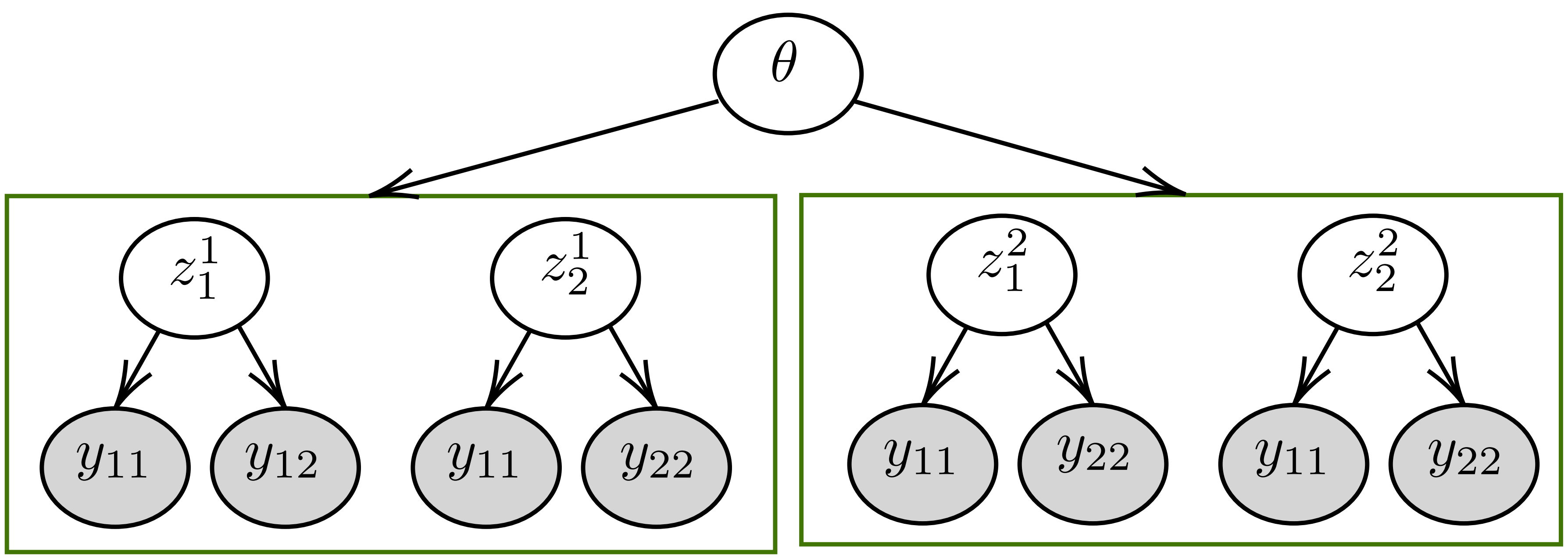
Subsampling: ✗

Enhanced bound: ✓

Local IW

Apply tightening to local variables separately.

$$\mathrm{LOCAL-ELBO} = \mathbb{E}\left[\log \frac{p(\theta)}{q(\theta)} + \sum_{i=1}^{M} \log \frac{1}{K} \sum_{k=1}^{K} \frac{p(z_i^k, y_i|\theta)}{q(z_i^k|\theta)}\right]$$



Subsampling: ✓

Enhanced bound: ✓

Local IW

Advantages:

- Applies Monte-Carlo in low-dimensional spaces.

- Compatible with subsampling.

Local IW

Advantages:

- Applies Monte-Carlo in low-dimensional spaces.
- Compatible with subsampling.

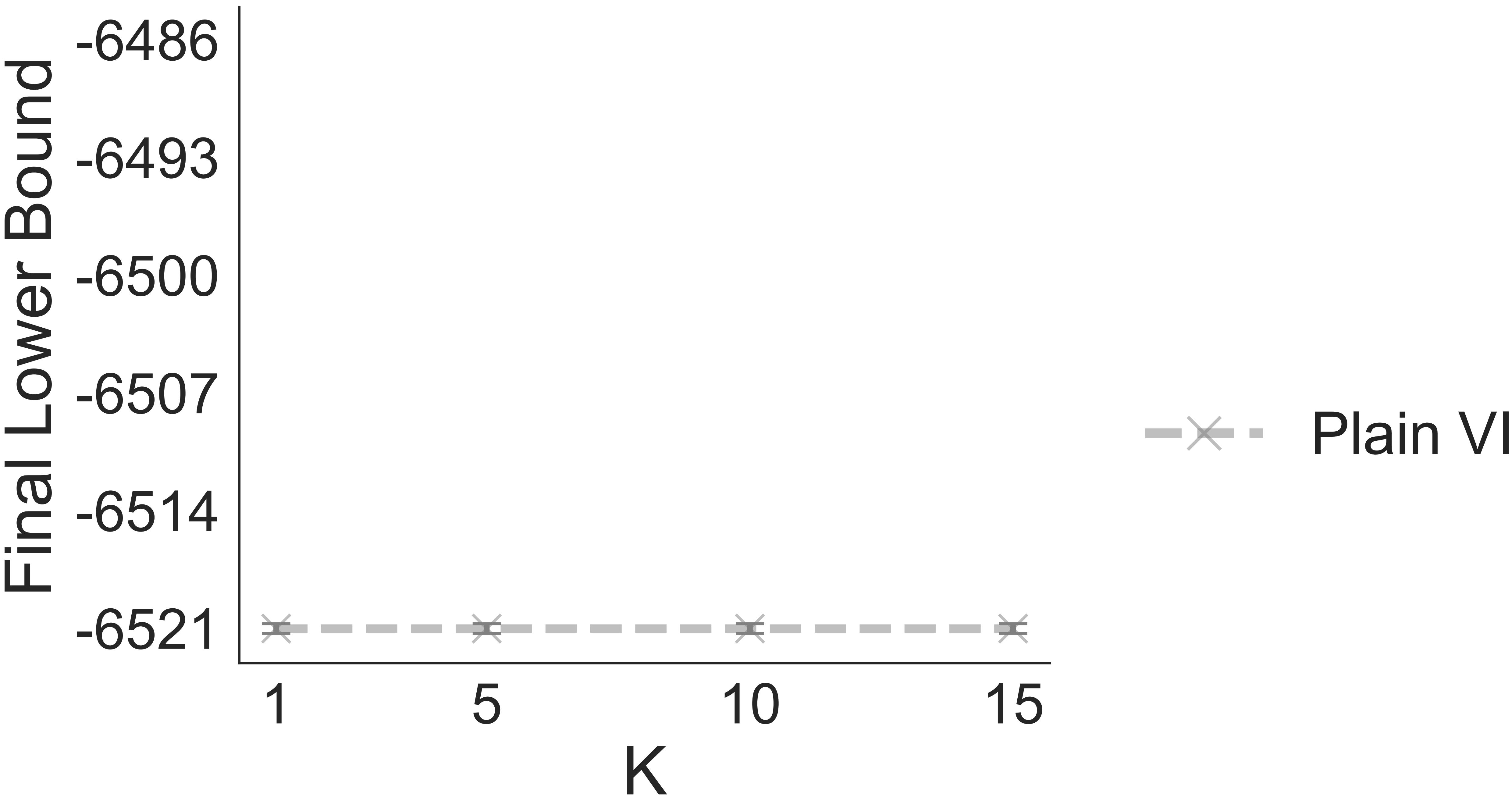
Price:

- For $K \to \infty$ variational gap is $KL(q(\theta) \| p(\theta|y))$

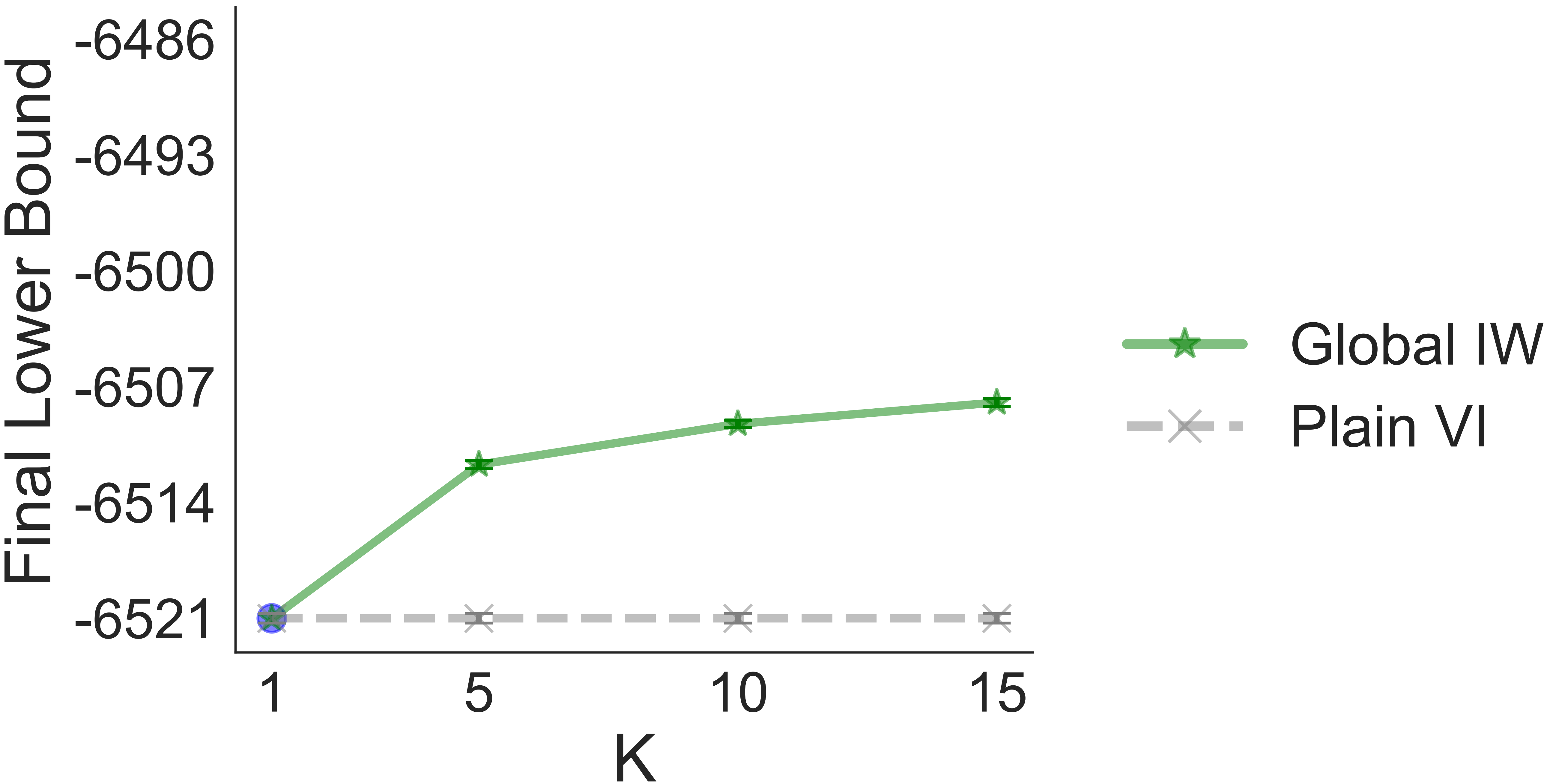
Global / Local AIS

(This is similar.)

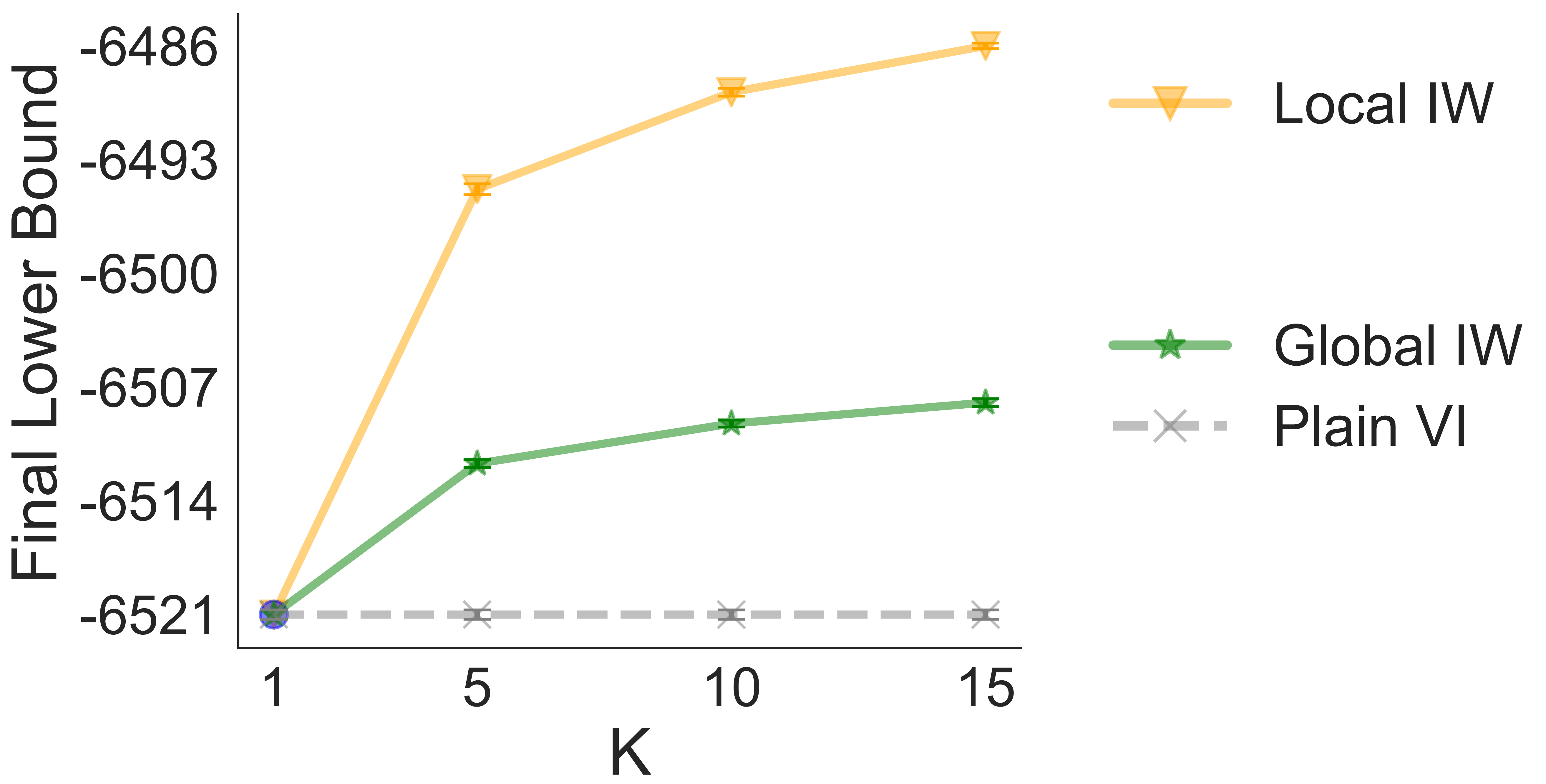
Experiments



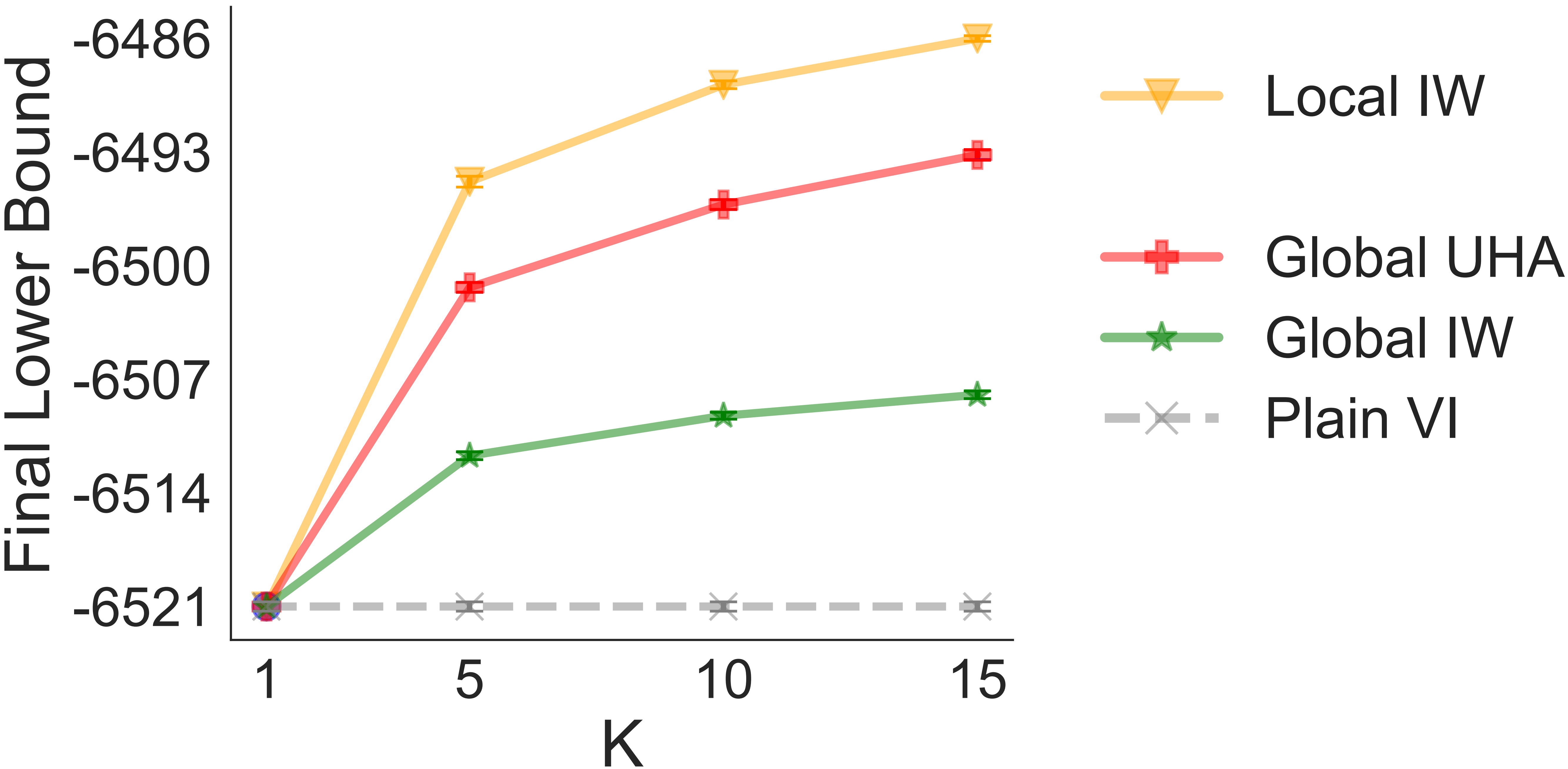
Experiments



Experiments



Experiments



Experiments