Structured Stochastic Gradient MCMC

Antony Alexos*, **Alex Boyd***, Stephan Mandt
University of California, Irvine

Bayesian Modeling

Posterior Distribution: $p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$

Methods: Analytically | MCMC | VI

Bayesian Modeling

Posterior Distribution:

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

Methods:

Analytically

MCMC

SGMCMC

VI

MFVI

Bayesian Modeling

Posterior Distribution: $p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$

Methods: Analytically | MCMC | VI

SGMCMC + MFVI

$\underbrace{\text{SGMCMC} + \text{MFVI}}_{\text{Structured SGMCMC}}$

Structured SGMCMC

True Posterior: $p(\theta|\mathcal{D})$

Approximate Posterior: $q(\theta) \coloneqq \prod_{i=1}^{M} q_i(\theta_i) = \exp\left\{-\sum_{i=1}^{M} U_i^{(S)}(\theta_i)\right\}$

$q_i^*(\theta_i) \propto \exp\left\{-U_i^{(S)}(\theta_i)\right\}$ where $U_i^{(S)}(\theta_i) = -\mathbb{E}_{\tilde{\theta}_{\neg i} \sim q_{\neg i}}\left[\log p\left(\theta_i, \tilde{\theta}_{\neg i}, \mathcal{D}\right)\right]$

Structured SGLD:

$$\theta^+ = \theta - \frac{\epsilon}{2} \nabla_\theta \widehat{U}^{(S)}\left(\theta; \widetilde{\mathcal{D}}\right) + \xi, \quad \xi \sim \mathcal{N}(0, \epsilon I)$$

Mini-batch $\widetilde{\mathcal{D}}$

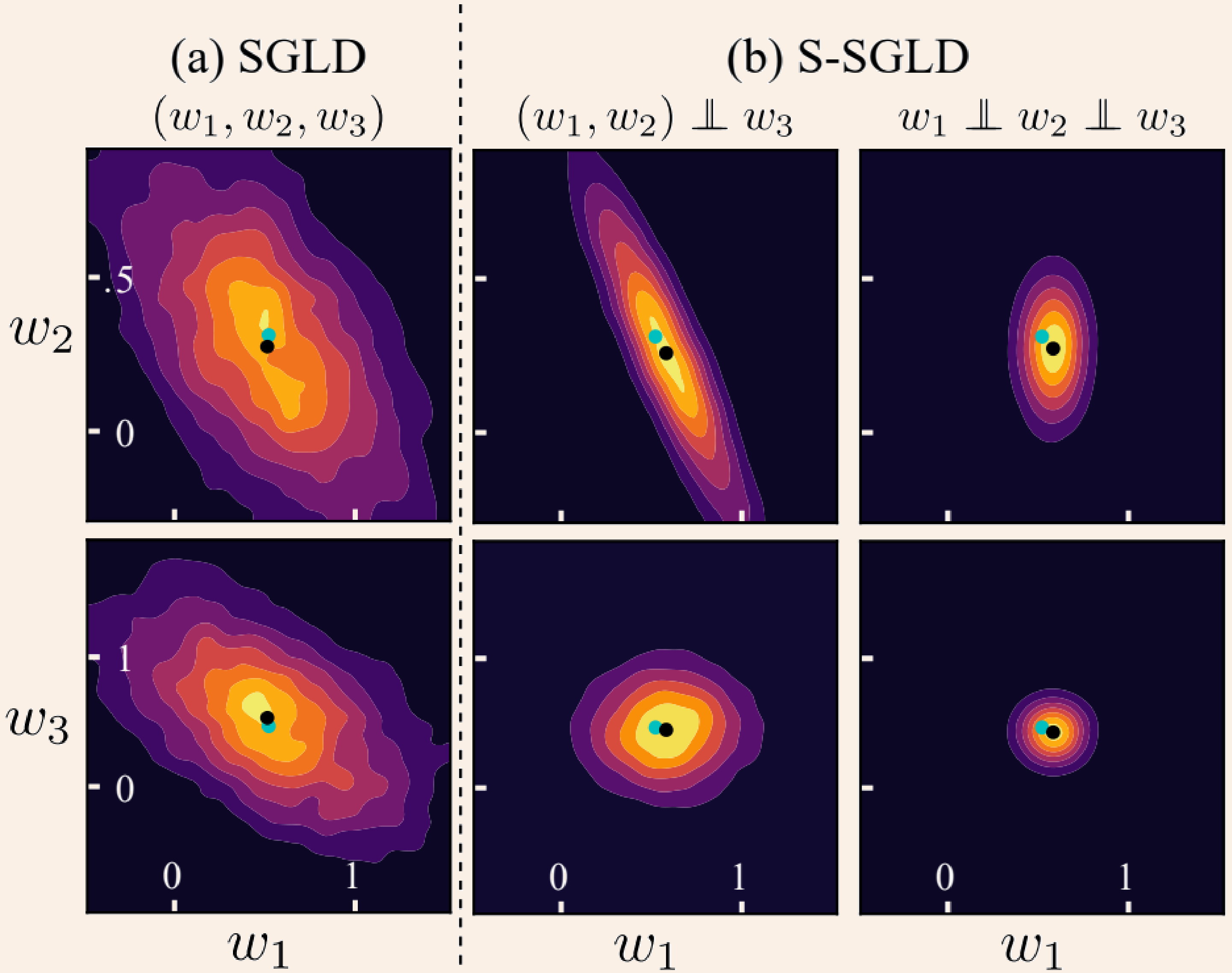
Approx. Structured Energy Function:

$$\widehat{U}^{(S)}\left(\theta; \widetilde{\mathcal{D}}\right) = \sum_{i=1}^{M} \mathbb{E}_{\tilde{\theta} \sim q}\, \widehat{U}\left(\{\theta_i, \tilde{\theta}_{\neg i}\}; \widetilde{\mathcal{D}}\right)$$

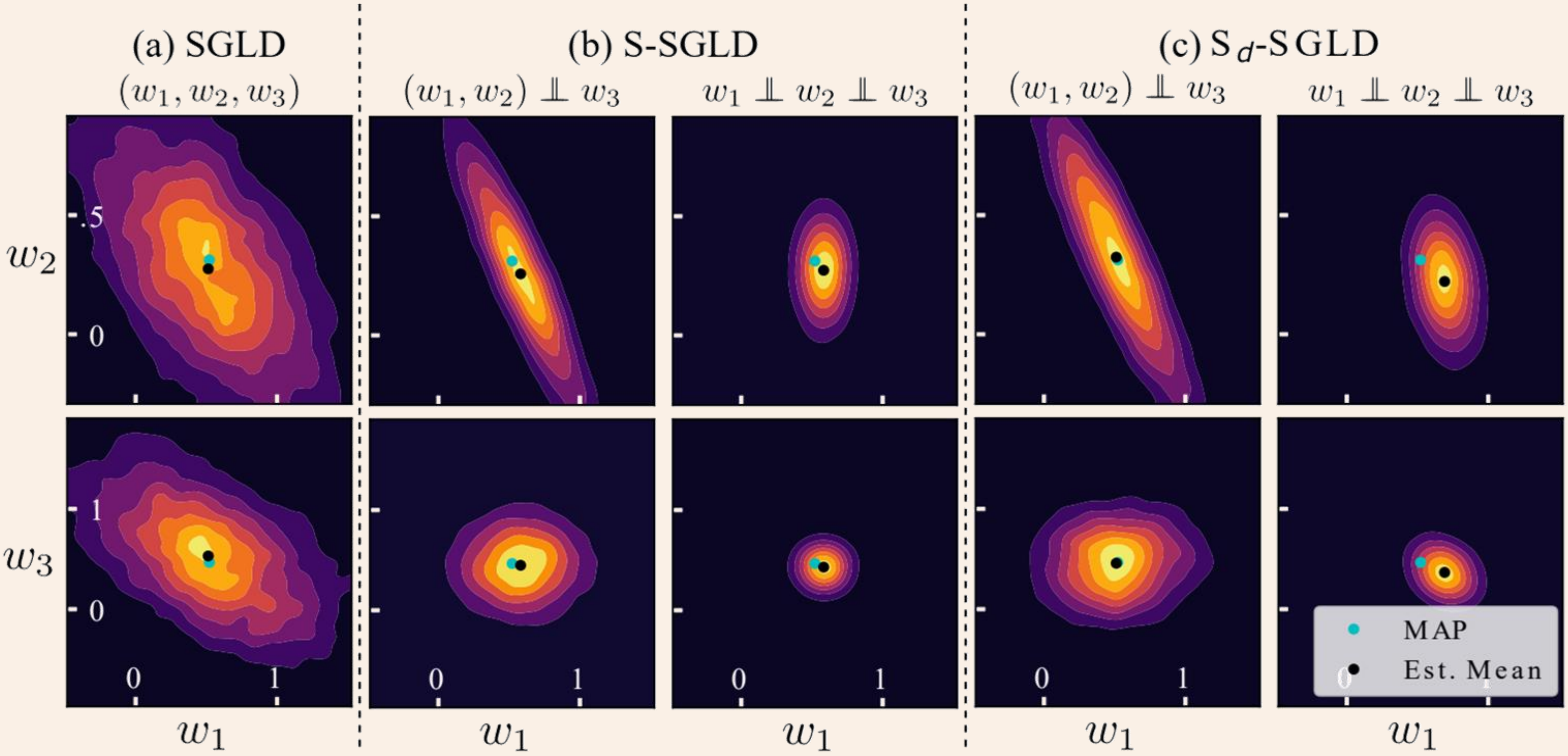
Approximate $\mathbb{E}_{\tilde{\theta} \sim q}$ with MC Samples:

$$\tilde{\theta}_i \sim \hat{q}_i(\theta_i) = \left\{\theta_i^{(1)}, \theta_i^{(2)}, \ldots\right\}$$

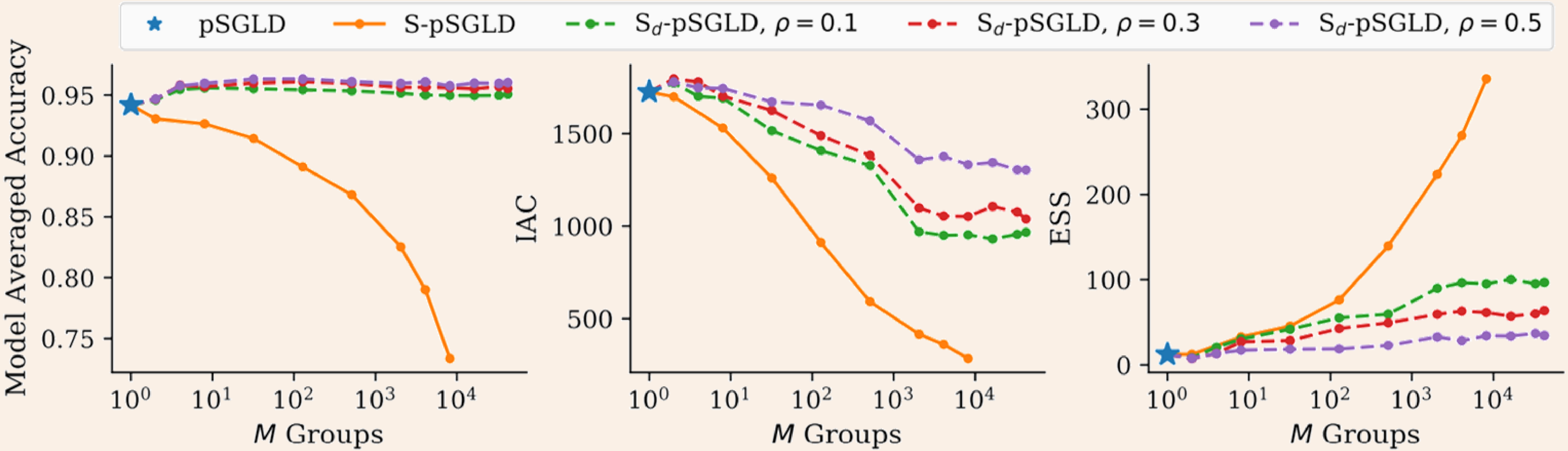
Posterior Visualization



Posterior Visualization



Method Comparison



Structured Stochastic Gradient MCMC

Antony Alexos*, **Alex Boyd***, Stephan Mandt
University of California, Irvine