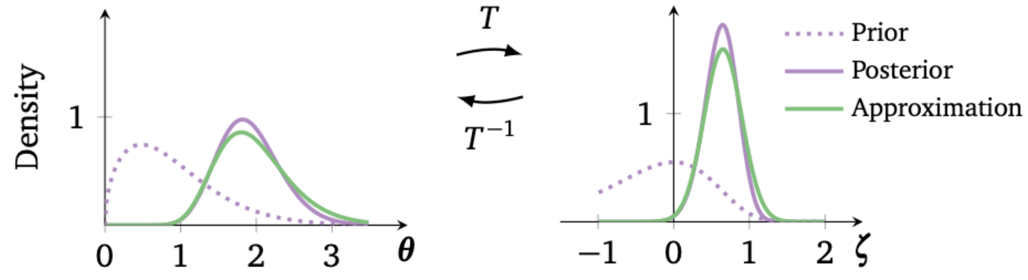
$$q(\boldsymbol{\theta}\,;\boldsymbol{\phi})$$

one-to-one differential transform

$$\zeta = T(\boldsymbol{\theta})$$

Mean-field Gaussian approximation

$$q(\boldsymbol{\zeta}\,;\boldsymbol{\phi}) = \mathcal{N}\left(\boldsymbol{\zeta}\,;\boldsymbol{\mu}, \operatorname{diag}(\boldsymbol{\sigma}^2)\right) = \prod_{k=1}^{K} \mathcal{N}\left(\zeta_k\,;\mu_k, \sigma_k^2\right)$$



(a) Latent variable space **(b)** Real coordinate space

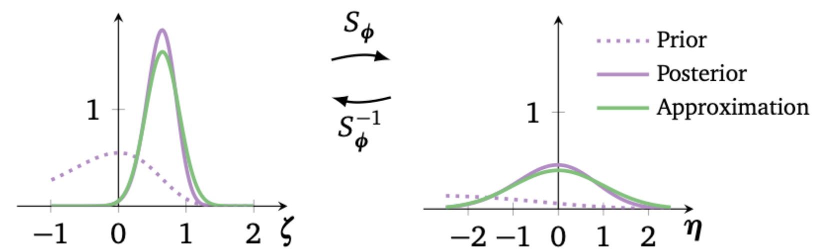
$$\boldsymbol{\phi}^* = \underset{\boldsymbol{\phi}\in\boldsymbol{\Phi}}{\arg\min} \operatorname{KL}\left(q(\boldsymbol{\theta}\,;\boldsymbol{\phi}) \,\|\, p(\boldsymbol{\theta} \mid \boldsymbol{x})\right)$$

Mnimize the KL divergence: maximize the ELBO

$$\boldsymbol{\phi}^* = \underset{\boldsymbol{\phi}\in\boldsymbol{\Phi}}{\arg\max}\, \mathscr{L}(\boldsymbol{\phi}) \quad \text{such that} \quad \operatorname{supp}(q(\boldsymbol{\theta}\,;\boldsymbol{\phi})) \subseteq \operatorname{supp}(p(\boldsymbol{\theta} \mid \boldsymbol{x}))$$

ELBO $\mathscr{L}(\boldsymbol{\phi}) = \mathbb{E}_{q(\boldsymbol{\theta})}\left[\log p(\boldsymbol{x}, \boldsymbol{\theta})\right] - \mathbb{E}_{q(\boldsymbol{\theta})}\left[\log q(\boldsymbol{\theta}\,;\boldsymbol{\phi})\right].$

$$\mathscr{L}(\boldsymbol{\phi}) = \mathbb{E}_{q(\boldsymbol{\zeta}\,;\boldsymbol{\phi})}\left[\log p\left(\boldsymbol{x}, T^{-1}(\boldsymbol{\zeta})\right) + \log\left|\det J_{T^{-1}}(\boldsymbol{\zeta})\right|\right] + \mathbb{H}\left[q(\boldsymbol{\zeta}\,;\boldsymbol{\phi})\right]$$



(a) Real coordinate space **(b)** Standardized space

Reparameterization

$$\boldsymbol{\phi}^* = \underset{\boldsymbol{\phi}}{\arg\max}\, \mathbb{E}_{\mathcal{N}(\boldsymbol{\eta}\,;0,\mathbf{I})}\left[\log p\left(\boldsymbol{x}, T^{-1}(S_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\eta}))\right) + \log\left|\det J_{T^{-1}}\left(S_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\eta})\right)\right|\right] + \mathbb{H}\left[q(\boldsymbol{\zeta}\,;\boldsymbol{\phi})\right]$$

---

step size

$$\rho_k^{(i)} = \eta \times i^{-1/2+\epsilon} \times \left(\tau + \sqrt{s_k^{(i)}}\right)^{-1}$$

adapts to the curvature of the ELBO

$$s_k^{(i)} = \alpha g_k^{2^{(i)}} + (1-\alpha) s_k^{(i-1)}$$