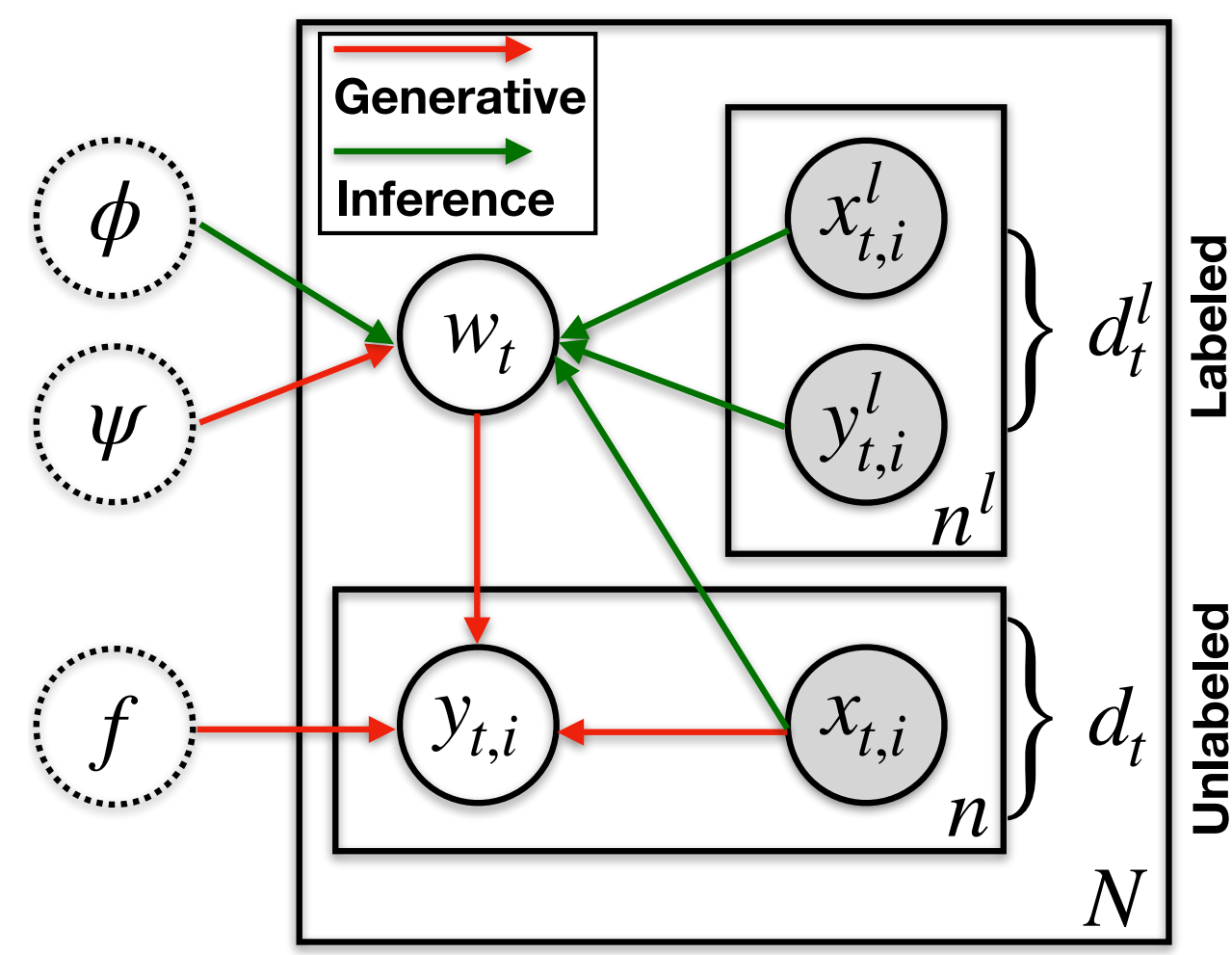
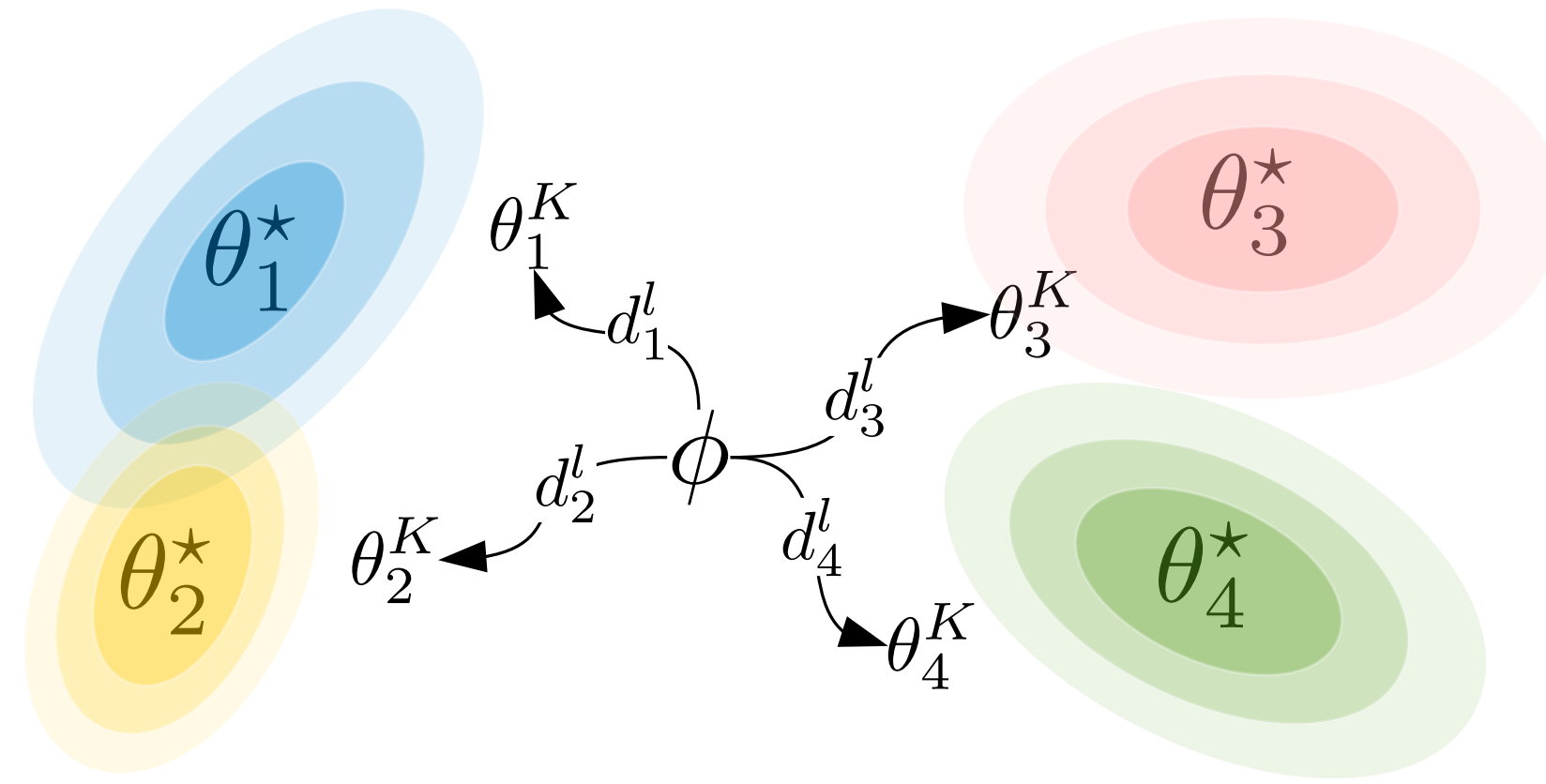


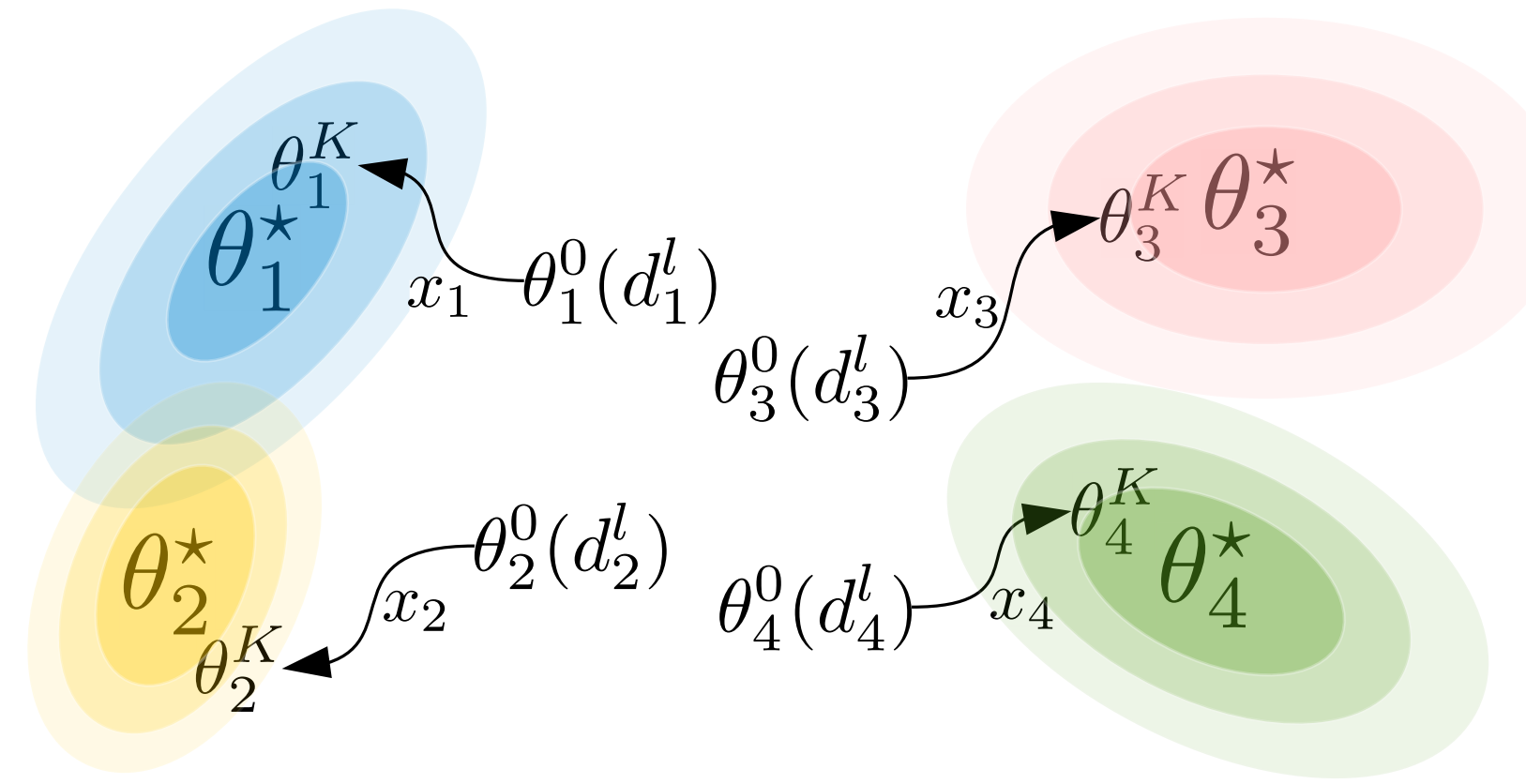
How can we make use of the unlabeled data (i.e., the query set) in meta-learning?



(a) Graphical model of EB



(b) MAML



(c) Our method (EBTML)

Contributions

- Transduction by synthetic gradients.
- SOTA results in Mini-ImageNet.
- Theoretical insights on empirical Bayes.

Figure 1. **A comparison between MAML and our method (EBTML)** is shown in (b) and (c). MAML is an inductive method since, for a task t , it first constructs a variational posterior $q_{\theta_t^k}$ (a Dirac delta distribution) as a function of the labeled set d_t^l , and then apply $q_{\theta_t^k}$ on the unlabeled set x_t ; while EBTML constructs a better variational posterior as a function of both d_t^l and x_t : it starts with an initialization $\theta_t^0(d_t^l)$ generated using the labeled set d_t^l , and then yields θ_t^K by running K synthetic gradient steps on the unlabeled set x_t .

From hierarchical Bayes to empirical Bayes

$$\text{HB : } p_f(\mathcal{D}) = \int_{\psi} \left[\prod_{t=1}^N \int_{w_t} p_f(d_t|w_t) p(w_t|\psi) \right] p(\psi)$$

$$\text{EB : } p_{\psi,f}(\mathcal{D}) = \prod_{t=1}^N \int_{w_t} p_f(d_t|w_t) p_{\psi}(w_t)$$

$$\text{Neg-loglik : } -\log p_f(d_t|w_t) = \sum_{i=1}^n \ell_t(\hat{y}_{t,i}(f(x_{t,i}), w_t), y_{t,i}) = L_t(w_t, d_t)$$

Variational inference for empirical Bayes

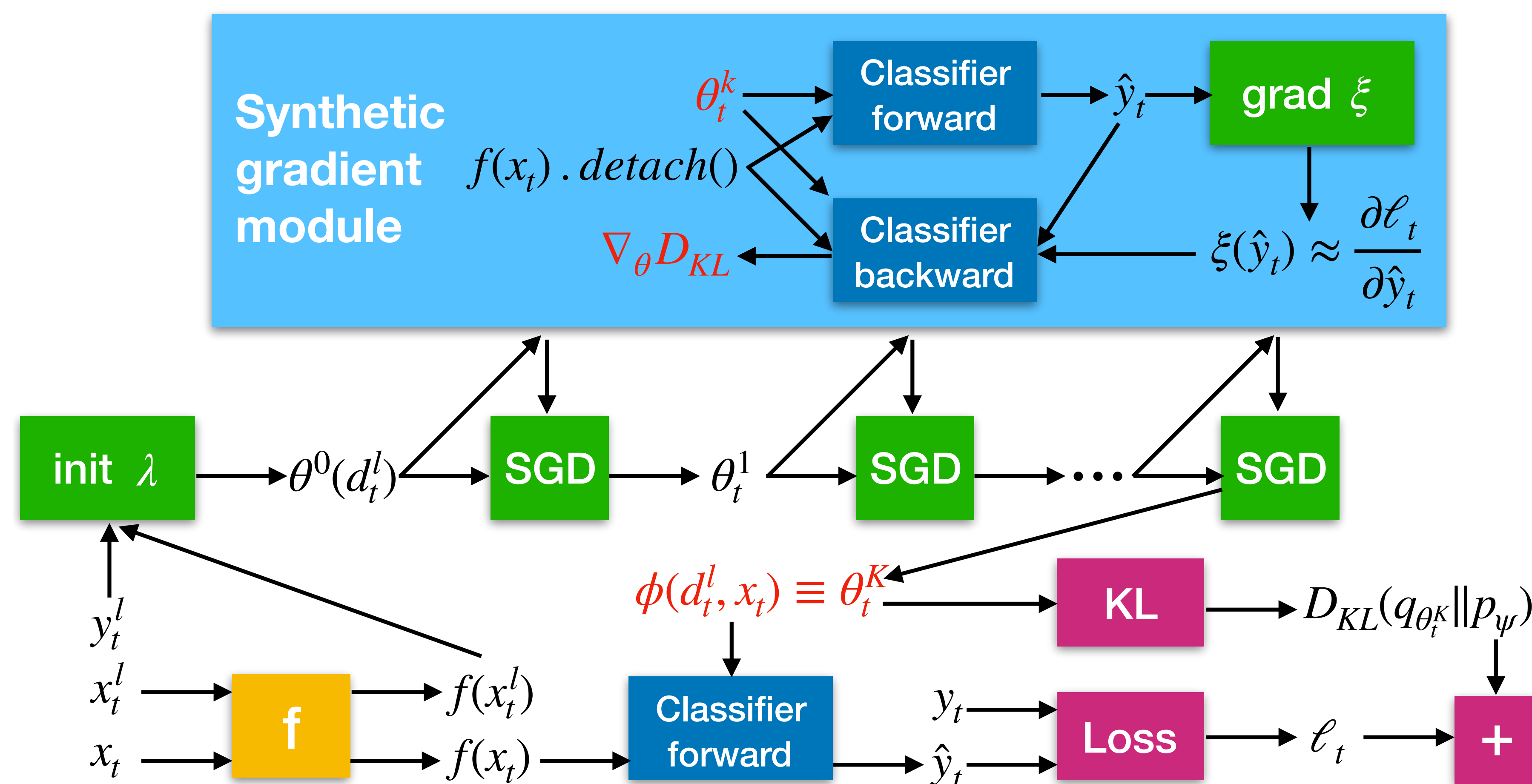
$$\text{Exact : } \min_{\psi,f} \min_{\theta_1, \dots, \theta_N} \sum_{t=1}^N D_{\text{KL}}(q_{\theta_t}(w_t) \parallel p_{\psi,f}(w_t|d_t))$$

$$\text{Inductive : } \min_{\psi,f} \min_{\phi} \sum_{t=1}^N D_{\text{KL}}(q_{\phi(d_t^l)}(w_t) \parallel p_{\psi,f}(w_t|d_t))$$

$$\text{Transductive : } \min_{\psi,f} \min_{\phi} \sum_{t=1}^N D_{\text{KL}}(q_{\phi(d_t^l, x_t)}(w_t) \parallel p_{\psi,f}(w_t|d_t))$$

Variational posterior via synthetic gradient [2] descent

$$\theta_t^{k+1} = \theta_t^k - \eta \left[\mathbb{E}_{\epsilon} \left[\sum_{i=1}^n \xi(\hat{y}_{t,i}) \frac{\partial \hat{y}_{t,i}}{\partial w_t} \frac{\partial w_t(\theta_t^k, \epsilon)}{\partial \theta_t} \right] + \nabla_{\theta_t} D_{\text{KL}}(q_{\theta_t^k} \parallel p_{\psi}) \right]$$



Few-shot classification on Mini-ImageNet

Method	FeatNet f	Mini-ImageNet, 5-way		CIFAR-FS, 5-way	
		1-shot	5-shot	1-shot	5-shot
MAML [1]	Conv-4-64	48.7±1.8%	63.1±0.9%	58.9±1.9%	71.5±1.0%
cc+rot [4]	Conv-4-64	54.8±0.4%	71.9±0.3%	63.5±0.3%	79.8±0.2%
EBTML $K=0$	Conv-4-64	50.0±0.4%	67.0±0.4%	59.2±0.5%	75.4±0.4%
EBTML $K=3$	Conv-4-64	58.0±0.6%	70.7±0.4%	68.7±0.6%	77.1±0.4%
cc+rot [4]	WRN-28-10	62.9±0.5%	79.9±0.3%	73.6±0.3%	86.1±0.2%
EBTML $K=0$	WRN-28-10	60.6±0.4%	77.5±0.3%	70.0±0.5%	83.5±0.4%
EBTML $K=1$	WRN-28-10	67.3±0.5%	78.8±0.4%	76.8±0.5%	84.9±0.4%
EBTML $K=3$	WRN-28-10	69.6±0.6%	78.9±0.4%	78.4±0.6%	85.3±0.4%
EBTML $K=5$	WRN-28-10	70.0±0.6%	79.2±0.4%	80.0±0.6%	85.3±0.4%

Link to information bottleneck [3]

Consider an abstract variational posterior $q(w|d, t)$ with inference & generative processes:

$$\text{Inference : } q(w, d, t) = q(t)q(d|t)q(w|d, t)$$

$$\text{Generative : } p(w, d, t) = p(d|w, t)p(w)q(t)$$

Theorem (generalization analysis of EB via IB)

If ℓ_t is σ -subgaussian under $q(w|t)q(z|t)$, then

$$\begin{aligned} \min_{p(w)} \mathbb{E}_{q(t)} \mathbb{E}_{q(d|t)} \left[D_{\text{KL}}(q(w|d, t) \parallel p(w|d, t)) \right] \\ \geq I_q(w; d|t) - \beta I_{q,p}(w; d|t) \text{ with } \beta = 1 \\ \geq \frac{n}{2\sigma^2} \text{gen}(q)^2 - \beta I_{q,p}(w; d|t), \end{aligned}$$

where I_q and $I_{q,p}$ are mutual information and cross mutual information respectively and

$$\text{gen}(q) = \mathbb{E}_{q(t)q(d|t)q(w|d, t)} \left[\mathbb{E}_{b \sim q(\cdot|t)} \log \frac{p(d|w, t)}{p(b|w, t)} \right]$$

Bibliography

- [1] Finn et al. Model-agnostic meta-learning for fast adaptation of deep networks. ICML 2017.
- [2] Jaderberg et al. Decoupled neural interfaces using synthetic gradients. ICML 2017.
- [3] Achille and Soatto. Emergence of invariance and disentangling in deep representations. JMLR 2018.
- [4] Gidaris et al. Boosting Few-Shot Visual Learning with Self-Supervision. ICCV 2019.

Paper and code

