M-step details

M-step details

$$\mathcal{L}(\theta, q) = \sum_i \sum_c q(t_i = c) \log \frac{p(x_i, t_i = c \mid \theta)}{q(t_i = c)}$$

M-step details

$$\mathcal{L}(\theta, q) = \sum_i \sum_c q(t_i = c) \log \frac{p(x_i, t_i = c \mid \theta)}{q(t_i = c)}$$

$$= \sum_i \sum_c q(t_i = c) \log p(x_i, t_i = c \mid \theta)$$

$$- \sum_i \sum_c q(t_i = c) \log q(t_i = c)$$

Const w.r.t. θ

M-step details

$$\mathcal{L}(\theta, q) = \sum_i \sum_c q(t_i = c) \log \frac{p(x_i, t_i = c \mid \theta)}{q(t_i = c)}$$

$$= \sum_i \sum_c q(t_i = c) \log p(x_i, t_i = c \mid \theta)$$

$$\cancel{- \sum_i \sum_c q(t_i = c) \log q(t_i = c)}$$

Const w.r.t. θ

M-step details

$$\mathcal{L}(\theta, q) = \sum_i \sum_c q(t_i = c) \log \frac{p(x_i, t_i = c \mid \theta)}{q(t_i = c)}$$

$$= \sum_i \sum_c q(t_i = c) \log p(x_i, t_i = c \mid \theta)$$

$$- \sum_i \sum_c q(t_i = c) \log q(t_i = c)$$

$$= \mathbb{E}_q \log p(X, T \mid \theta) + \text{const}$$

M-step details

$$\mathcal{L}(\theta, q) = \sum_i \sum_c q(t_i = c) \log \frac{p(x_i, t_i = c \mid \theta)}{q(t_i = c)}$$

$$= \sum_i \sum_c q(t_i = c) \log p(x_i, t_i = c \mid \theta)$$

$$- \sum_i \sum_c q(t_i = c) \log q(t_i = c)$$

$$= \boxed{\mathbb{E}_q \log p(X, T \mid \theta)} + \text{const}$$

(Usually) concave function w.r.t. θ, easy to optimize

Expectation Maximization algorithm

For k = 1, …

E-step

$$q^{k+1} = \arg\min_q \mathcal{KL}\left[q(T) \parallel p(T \mid X, \theta^k)\right]$$

$$\Leftrightarrow$$

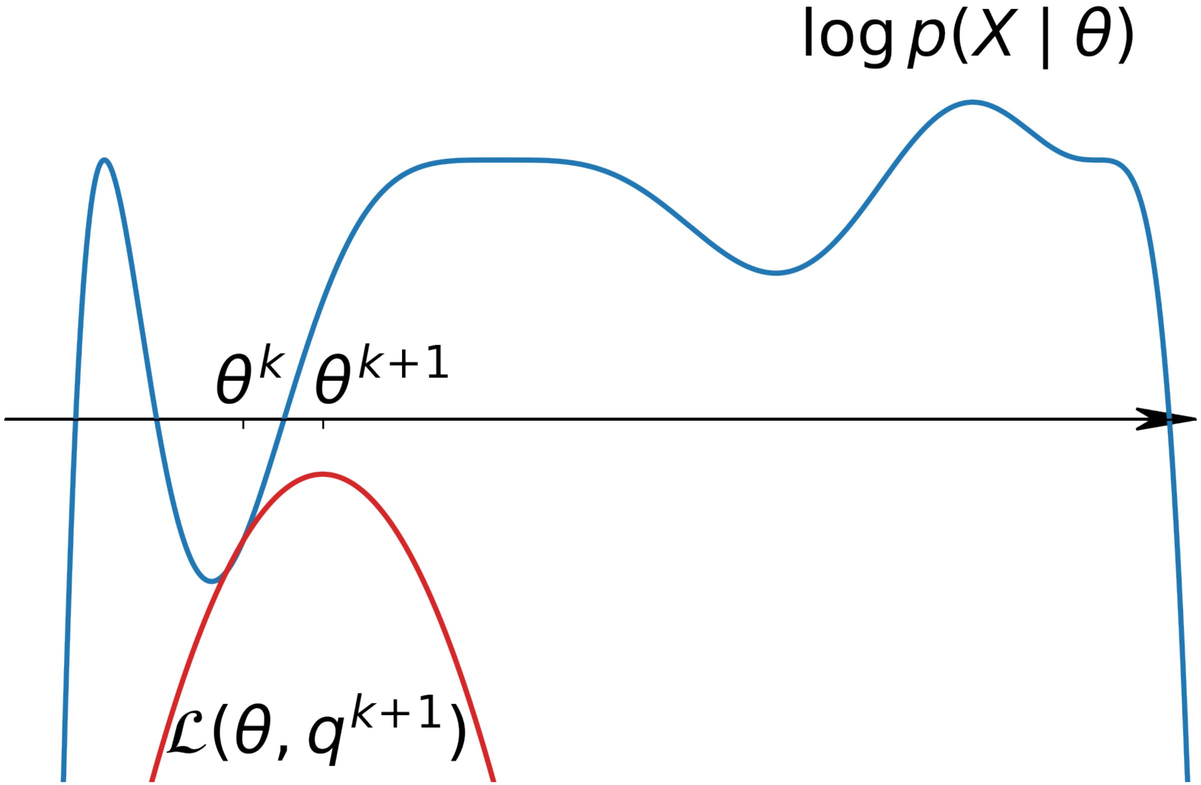
$$q^{k+1}(t_i) = p(t_i \mid x_i, \theta^k)$$

M-step

$$\theta^{k+1} = \arg\max_\theta \mathbb{E}_{q^{k+1}} \log p(X, T \mid \theta)$$

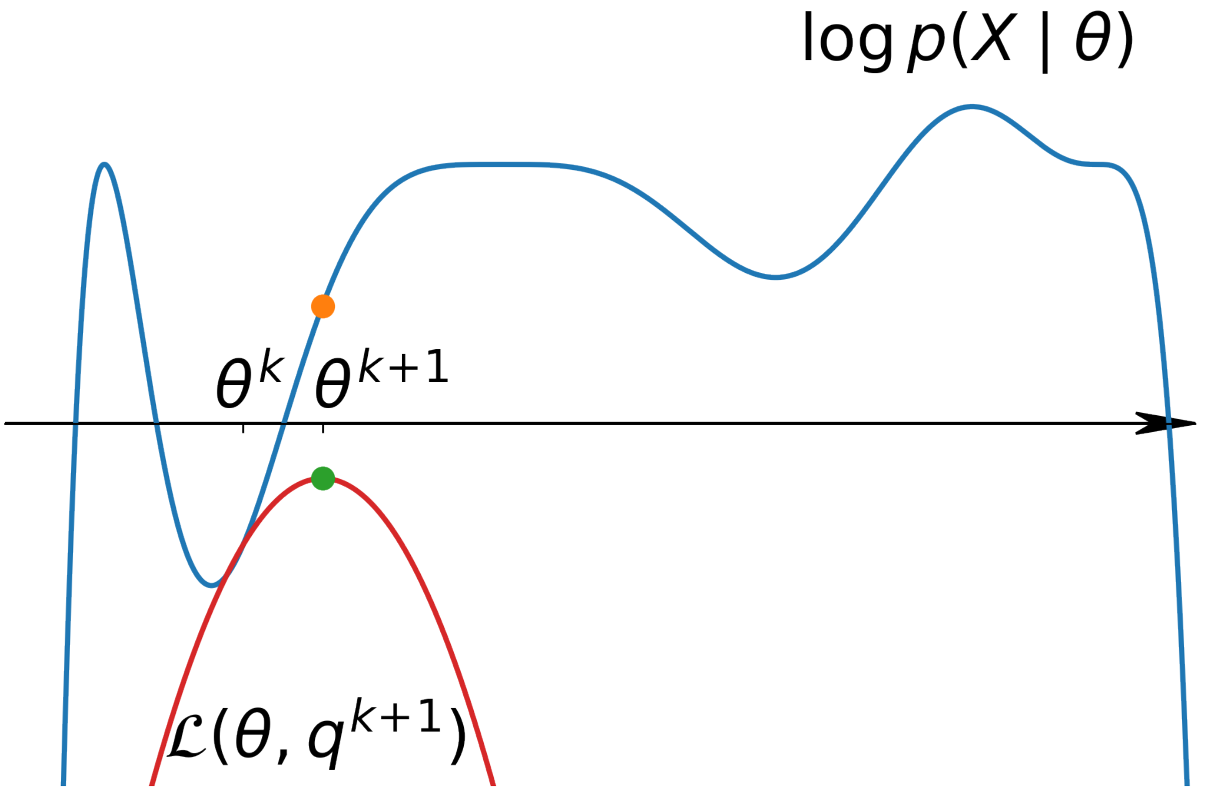
Convergence guaranties

Convergence guaranties



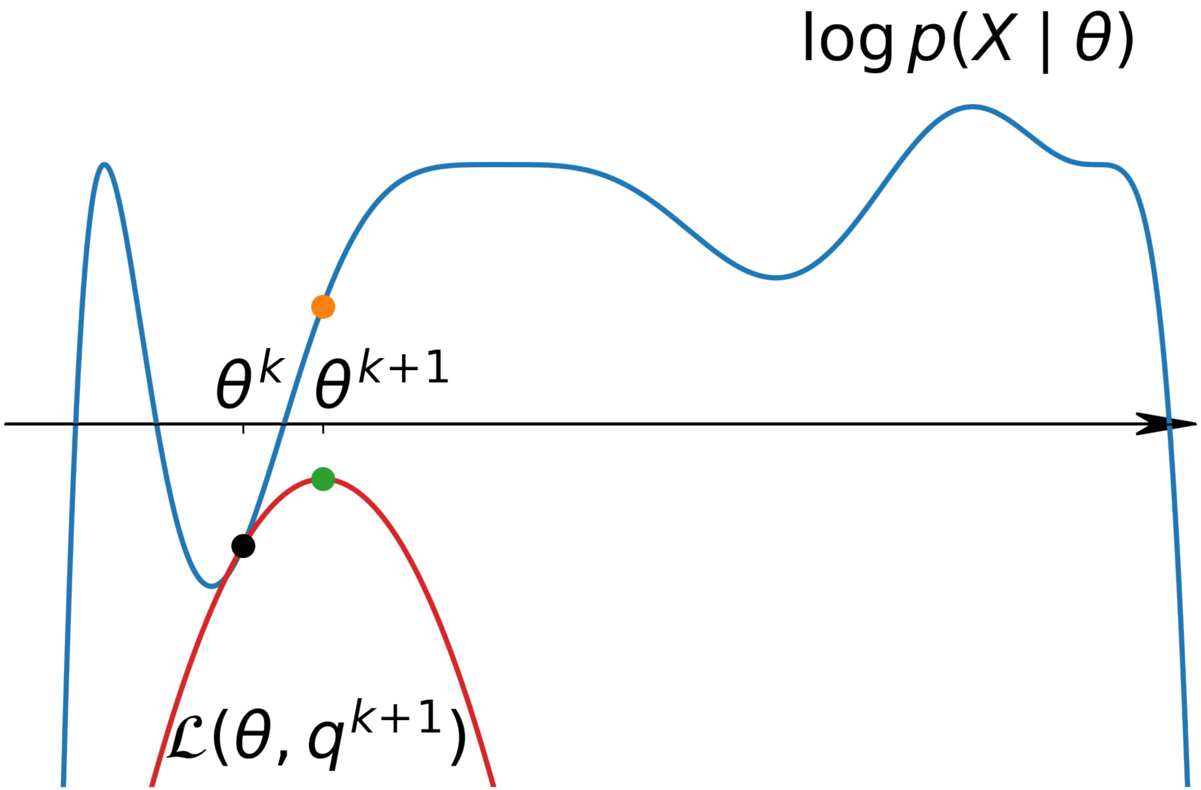
Convergence guaranties

$$\log p(X \mid \theta^{k+1}) \geq \mathcal{L}(\theta^{k+1}, q^{k+1})$$



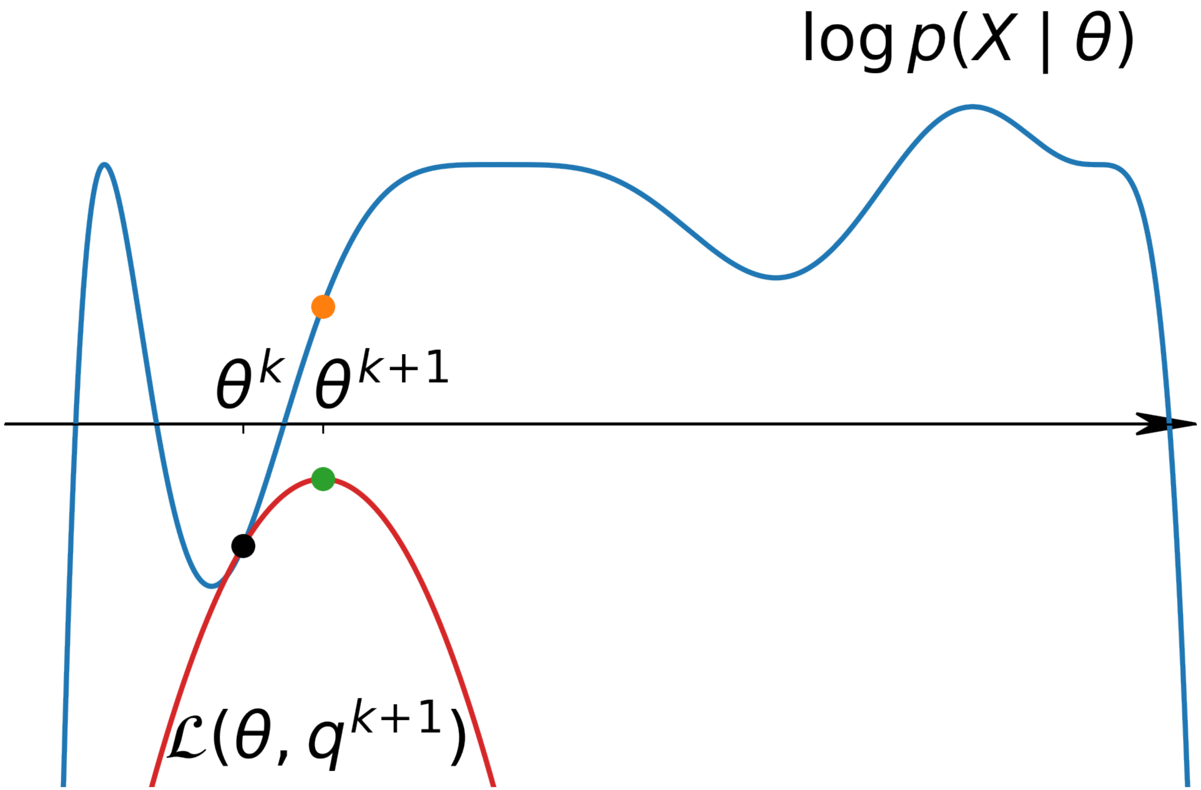
Convergence guaranties

$$\log p(X \mid \theta^{k+1}) \geq \mathcal{L}(\theta^{k+1}, q^{k+1}) \geq \mathcal{L}(\theta^{k}, q^{k+1})$$



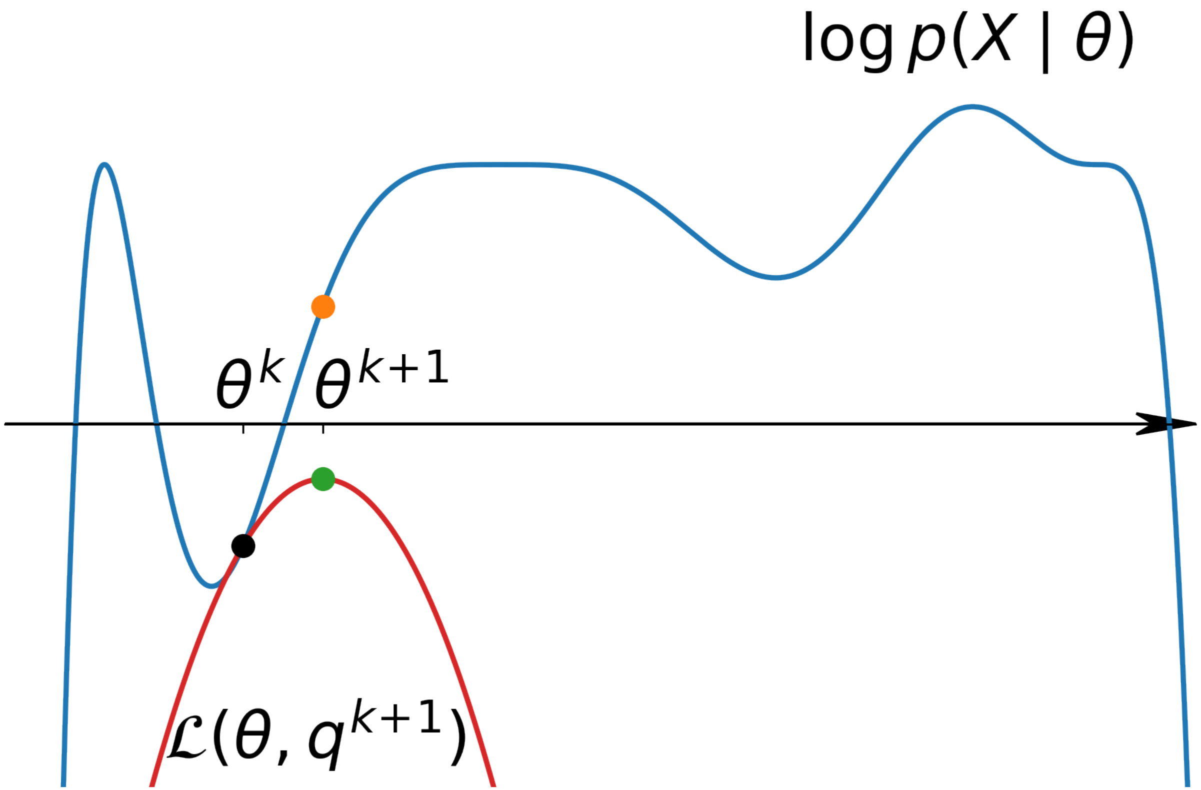
Convergence guaranties

$$\log p(X \mid \theta^{k+1}) \geq \mathcal{L}(\theta^{k+1}, q^{k+1}) \geq \mathcal{L}(\theta^{k}, q^{k+1}) = \log p(X \mid \theta^{k})$$



Convergence guaranties

$$\log p(X \mid \theta^{k+1}) \geq \log p(X \mid \theta^k)$$



Convergence guaranties

$$\log p(X \mid \theta^{k+1}) \geq \log p(X \mid \theta^k)$$

- On each iteration EM doesn't decrease the objective (good for debugging!)
- Guarantied to converge to a local maximum (or saddle point)