Thus one function to contemplate comes to the front:

$$Q(\vec{\theta};\vec{\theta^i}) = E_{D_b}[\ln p(D_g, D_b; \vec{\theta})|D_g; \vec{\theta^i}] \tag{1}$$

- $Q(\vec{\theta};\theta^i)$ is a function of $\vec{\theta}$ and $\vec{\theta^i}$
- $E_{D_b}[\ln p(D_g, D_b; \vec{\theta})|D_g; \vec{\theta^i}]$ is the expected value is over the missing features. The expected value hinges on $\vec{\theta^i}$ are the true parameters.
- $\vec{\theta^i}$ is the current (best) estimate for the full distribution;
- $\vec{\theta}$ is a candidate vector for an improved estimate
- D_b gets marginalized with respect to $\vec{\theta^i}$.
- The goal of the EM algorithm is select from the candidate $\vec{\theta}$ from a set of $\vec{\theta}$s, and iterate it to $\vec{\theta}^{i+1}$ which yields the greatest $Q(\vec{\theta};\vec{\theta^i})$