Guide to slides of EM Convergence Presentation

Chen, Yu

December 2020

1 Important Concepts

1.1 Closed Point-to-set Map

- **Point-to-set map** (p2s map) M: map from pints of a set X to subsets of X
- M is **closed** at x^*: $x_k \in X$, $\lim\limits_{k\to\infty} x_k = x^*$ and $\lim\limits_{k\to\infty} y_k = y^*$, $y_k \in M(x_k)$ imply $y^* \in M(x^*)$

1.2 Curved Exponential Family

X is a random vector with p.d.f. $f(x|\phi)$ from a probability space X.

$$f(x|\phi) = A(\phi)\exp\left(\sum_{i=1}^{k} T_i(x)\eta_i(\phi)\right)h(x)$$

Where $T_i(x)$ is a real valued statistics, $\eta_i(\phi)$ is a real valued function on the parameter space $\Omega \subseteq R^q$, and $q < k \in N$.
If $cov_\phi(\vec{T})$ ($\vec{T} = [T_1, T_2, \cdot, T_k]$) is positive definite, then X belongs to the curved exponential family.

1.3 Three Important Sets

$$\varphi(a) = \{\phi \in \varphi : L(\phi) \equiv a\}$$
$$M(a) = \{\phi \in M : L(\phi) \equiv a\}$$
$$\psi(L) = \{\phi \in \Omega : L(\phi) \equiv L\}$$

where M: local maxima in the interior of Ω
φ: stationary points in the interior of Ω
Ω: parameter space

2 Form of GEM

$$\max_{\phi'} L(\phi') = Q(\phi'|\phi) - H(\phi'|\phi) = E\{log(f(x|\phi'))|y,\phi\} - E\{log(k(x|y,\phi'))|y,\phi\} \tag{1}$$

- E-STEP Determine $Q(\phi|\phi_p)$ for current ϕ_p
- M-STEP $\phi_p \to \phi_{p+1} \in M(\phi_p) = \{\phi; \phi =_{\phi\in\Omega} Q(\phi;\phi_p)\}$

3 Assumptions of GEM

- 1) $\Omega \subseteq R^r$
- 2) $\Omega_{\phi_0} = \{\phi \in \Omega : L(\phi) \geq L(\phi_0)\}$ is compact for any $L(\phi_0) > -\infty$
- 3) $L(\phi)$ is continuous in Ω and differentiable in the interior of Ω
- 4) $\{L(\phi_p)\}_{p\geq 0}$ is bounded above for any $\phi_0 \in \Omega$
- 5) ϕ_p is in the interior of Ω, $int(\Omega)$
- 6) ϕ_p converges to some $\phi^* \in int(\Omega)$ such that the Hessian matrices $\nabla^2 Q(\phi^*|\phi^*)$ and $\nabla^2 H(\phi^*|\phi^*)$ exist at the first ϕ^*, and $\nabla^2 Q(\phi'|\phi)$ is continuous in (ϕ', ϕ)

4 Important Properties of GEM

- Any EM sequence $\{\phi_p\}$ increases the likelihood and $L(\phi_p)$, if bounded above, converges to some L^*
- Property of point-to-set map:
$$Q(\phi^{'}|\phi) \geq Q(\phi|\phi) \;\forall\; \phi^{'} \in M(\phi) \tag{2}$$
- According to Theorem 1 of DLR paper:
$$H(\phi|\phi) \geq H(\phi^{'}|\phi) \;\forall\; \phi^{'} \in \Omega \tag{3}$$
- Lemma 1 of DLR: for any sequence $\{\phi_p\}$ from a GEM algorithm:
$$L(\phi_{p+1}) \;\geq\; L(\phi_p) \tag{4}$$

5 Guide to the Theorems

Global Convergence Theorem from linear programming is the root of the first three theorems. Theorem 1 to Theorem 3 are dedicated to answering the first question: $L(\phi_p)$ Converges to Global Maximum, Local Maximum or Stationary Value? Theorem 4 to Theorem 7 answer the second question: Convergence of an GEM sequence $\{\phi_p\}$.

Specifically, Theorem 1 is a speciality or application of the Global Convergence Theorem in EM algorithm. Theorem 1 introduces 2 conditions:
1) M is a closed p2s map over φ^C (or M^C)
2) $L(\phi_{p+1}) > L(\phi_p) \;\forall \phi_p \notin \varphi$ (or M)
which are used as the conditions in Theorem 4 and Theorem 5.

Theorem 2 propose the continuity: $Q(\psi|\phi)$ is continuous in ψ and ϕ, which implies the 2 conditions in Theorem 1, but this theorem can only apply to the situation of stationary point not the situation of local maximum.

Theorem 3 add an extra condition: $sup_{\phi^{'} \in \Omega} Q(\phi^{'}|\phi) > Q(\phi|\phi) \;\forall\; \phi \in \varphi \backslash M$ to Theorem 2 to make Theorem 2 to be able to apply to the maximum case.

Theorem 4 inherits the two conditions from Theorem 1 and add another condition: $\varphi(L^*) = \{\phi^*\}$ (or $M(L^*) = \{\phi^*\}$) where $L^* = \lim_{p\to\infty} L(\phi_p)$ to answer the second question.

Theorem 5 relaxes the condition $\varphi(L^*) = \{\phi^*\}$ (or $M(L^*) = \{\phi^*\}$) where $L^* = \lim_{p\to\infty} L(\phi_p)$ of Theorem 4 using $\lim_{p\to\infty} ||\phi_{p+1} - \phi_p|| = 0$, such that answers the question of regarding the convergence of $\{\phi_p\}$.

Theorem 6, unlike Theorem 4 or 5, gives up the two conditions from Theorem 1, and replaces the two by: $\nabla Q(\phi_{p+1}|\phi_p) = 0$ and $\nabla Q(\phi^{'}|\phi)$ is continuous in $\phi^{'}$ and ϕ, and a stronger condition: (a) $\psi(L^*) = \{\phi^*\}$, or (b) $\lim_{p\to\infty} ||\phi_{p+1} - \phi_p|| = 0$ and $\psi(L^*)$ is discrete.

Theorem 7 is a corollary of Theorem 6, and points out under which conditions the sequence $\{\phi_p\}$ converges to the only maxima.

The most applicable conclusions are Theorem 2 and 7.