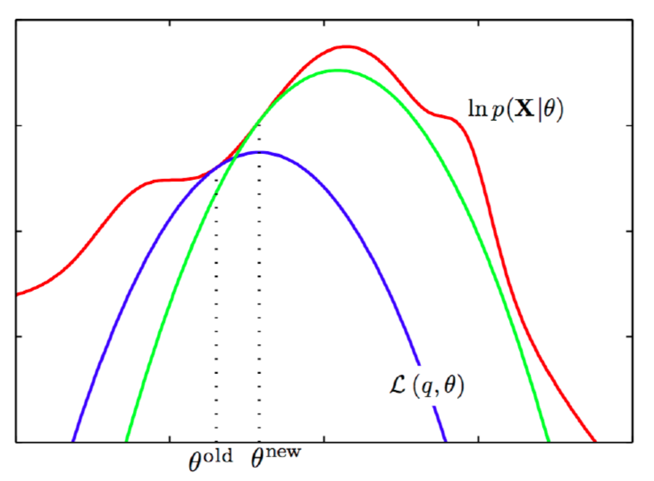
The Convergence Property of EM

Yu Chen

14th December 2020

Technical University of Munich

Intuition of EM



Outline

1. Introduction to the General EM
2. $L(\phi_p)$ Converges to Global Maximum, Local Maximum or Stationary Value?
3. Convergence of an GEM sequence $\{\phi_p\}$
4. Summary

Introduction to the General EM

General Form of EM Algorithm – 1

We have 2 sample spaces $\mathbb{X}$ and $\mathbb{Y}$ with corresponding p.d.f. $f(\boldsymbol{x}|\phi)$ and $g(\boldsymbol{y}|\phi)$ satisfying the following relationship:

$$g(\boldsymbol{y}|\phi) = \int_{\{\boldsymbol{x}:\boldsymbol{y}(\boldsymbol{x})=\boldsymbol{y}\}} f(\boldsymbol{x}|\phi)d\boldsymbol{x}$$

where $\boldsymbol{y} = \boldsymbol{y}(\boldsymbol{x})$ is the observed incomplete data in $\mathbb{Y}$ and $\phi \in \Omega$ the parameter space.

NOTE: a modern expression:

$$g(\boldsymbol{y}|\phi) = \int f(\boldsymbol{x}, \boldsymbol{y}|\phi)d\boldsymbol{x}$$

General Form of EM Algorithm – 2

Let $k(\boldsymbol{x}|\boldsymbol{y},\phi) = \frac{f(\boldsymbol{x}|\phi)}{g(\boldsymbol{y}|\phi)}$, then the log likelihood $L(\phi^{'})$ is:

$$\begin{aligned} L(\phi^{'}) &= log(g(\boldsymbol{y}|\phi^{'})) \\ &= E_{\boldsymbol{x}\sim k(\boldsymbol{x}|\boldsymbol{y},\phi)}[log(g(\boldsymbol{y}|\phi^{'}))] \\ &= E_{\boldsymbol{x}\sim k(\boldsymbol{x}|\boldsymbol{y},\phi)}[log(\frac{f(\boldsymbol{x}|\phi^{'})}{k(\boldsymbol{x}|\boldsymbol{y},\phi^{'})})] \\ &= E_{\boldsymbol{x}\sim k(\boldsymbol{x}|\boldsymbol{y},\phi)}[log(f(\boldsymbol{x}|\phi^{'})) - log(k(\boldsymbol{x}|\boldsymbol{y},\phi^{'}))] \\ &= E\{log(f(\boldsymbol{x}|\phi^{'}))|\boldsymbol{y},\phi\} - E\{log(k(\boldsymbol{x}|\boldsymbol{y},\phi^{'}))|\boldsymbol{y},\phi\} \\ &= Q(\phi^{'}|\phi) - H(\phi^{'}|\phi) \end{aligned} \tag{1}$$

General Form of EM Algorithm – 3

We are interested in $\max_{\phi'} L(\phi') = Q(\phi'|\phi) - H(\phi'|\phi)$.
Solve using EM iteration:

$$\phi_p \to \phi_{p+1} \in M(\phi_p) = \{\phi; \phi = \arg\max_{\phi \in \Omega} Q(\phi; \phi_p)\}$$

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

Generalized EM (GEM) Algorithm

Dempster, Laird and Rubin (1997) defined the GEM algorithm in their paper (abbreviated DLR paper[2]) as an iterative scheme:

$$\phi_{p+1} \in M(\phi_p) = \{\phi; \phi = \arg\max_{\phi \in \Omega} Q(\phi; \phi_p)\}$$

where $\phi \to M(\phi)$ is a point-to-set map such that:

$$Q(\phi^{'}|\phi) \geq Q(\phi|\phi) \ \forall \ \phi^{'} \in M(\phi) \tag{2}$$

According to Theorem 1 and Lemma 1 of DLR:

$$H(\phi|\phi) \geq H(\phi^{'}|\phi) \ \forall \ \phi^{'} \in \Omega \tag{3}$$

For any sequence $\{\phi_p\}$ from a GEM algorithm:

$$L(\phi_{p+1}) \ \geq \ L(\phi_p) \tag{4}$$

$L(\phi_p)$ Converges to Global Maximum, Local Maximum or Stationary Value?

Assumptions

- 1) $\Omega \subseteq \mathbb{R}^r$
- 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 $(\phi^{'}, \phi)$

Hessian Matrix of log-likelihood at ϕ^*

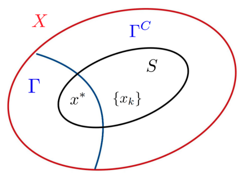
- $\nabla^2 Q(\phi^*|\phi^*)$ is non-positive definite (n.p.d.);
- $-\nabla^2 H(\phi^*|\phi^*)$ is non-negative definite (n.n.d.) (Lemma 2 of DLR);

So, $\nabla^2 L(\phi^*) = \nabla^2 Q(\phi^*|\phi^*) - \nabla^2 H(\phi^*|\phi^*)$ may not be n.p.d. i.e. $L(\phi^*)$ may not be a local maximum.

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^*: $x_k \in X$, $\lim_{k\to\infty} x_k = x^*$ and $\lim_{k\to\infty} y_k = y^*$, $y_k \in M(x_k)$ imply $y^* \in M(x^*)$

Global Convergence Theorem, Zangwill 1969



- $\{x_k\}_{k=0}^{\infty}$ is generated by $x_{k+1} \in M(x_k)$ where M is p2s map on set X
- Solution set $\Gamma \subset X$
  - 1) $x_k \in S$ where compact set $S \subset X$
  - 2) M is closed over Γ^C

  Then all limit points of $\{x_k\}$ are in Γ
  - 3) function α is continuous on X such that

  a) if $x \in \Gamma^C$, then $\alpha(y) > \alpha(x)\ \forall\ y \in M(x)$;

  b) if $x \in \Gamma$, then $\alpha(y) \geq \alpha(x)\ \forall\ y \in M(x)$.

  Then $\alpha(x_k)$ converges monotonically to $\alpha(x^*)$ for some $x^* \in \Gamma$

Global Convergence Theorem on GEM – 1

- M be the p2s map in a GEM iteration
- $\alpha(x)$ be the log-likelihood L
- Let the solution set Γ be:
  - $\mathscr{M}$: local maxima in the interior of Ω
  - φ: stationary points in the interior of Ω

Theorem 1

ϕ_p is a GEM sequence generated by $\phi_{p+1} \in M(\phi_p)$, and suppose:
1) M is a closed p2s map over φ^C (or $\mathscr{M}^C$)
2) $L(\phi_{p+1}) > L(\phi_p)\ \forall \phi_p \notin \varphi$ (or $\mathscr{M}$)
Then, all limit points of $\{\phi_p\}$ are stationary (or local maxima) of L, and $L(\phi_p)$ converges monotonically to $L^* = L(\phi^*)$

NOTE: if $Q(\psi|\phi)$ is continuous in ψ and ϕ, then condition 1) in Theorem 1 satisfies.

Global Convergence Theorem on GEM – 2

Theorem 2

If $Q(\psi|\phi)$ is continuous in ψ and ϕ,
then all limit points of $\{\phi_p\}$ in a GEM are stationary points of L, and $L(\phi_p)$ converges monotonically to $L^* = L(\phi^*)$ for some point ϕ^*

Proof

Condition 1) of Theorem 1 has held, only need to prove condition 2).
$\because \mathrm{H}(\phi|\phi) \geq H(\phi^{'}|\phi)\ \forall \phi^{'} \in \Omega$ (Theorem 1 of DLR)
$\therefore \nabla H(\phi_p|\phi_p) = 0$
$\therefore \nabla L(\phi_p) = \nabla Q(\phi_p|\phi_p) \neq 0\ \forall\ \phi_p \notin \varphi$
$\because \mathrm{Q}(\phi_{p+1}|\phi_p) > Q(\phi_p|\phi_p)$ and $H(\phi_p|\phi_p) \geq H(\phi_{p+1}|\phi_p)$
$\therefore \mathrm{L}(\phi_{p+1}) > L(\phi_p)$ ■
NOTE: Theorem 2 DOES NOT apply to $\mathscr{M}$ (consider some $\phi_p \in \varphi$ whereas $\phi_p \notin \mathscr{M}$)

Global Convergence Theorem on GEM – 2

Curved exponential family is a class of densities satisfying the continuity in Theorem 2.

Curved Exponential Family

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

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

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

Global Convergence Theorem on GEM – 3

Theorem 3

If $Q(\psi|\phi)$ is continuous in ψ and ϕ, and
$sup_{\phi' \in \Omega} Q(\phi^{'}|\phi) > Q(\phi|\phi) \ \forall \ \phi \in \varphi \backslash \mathscr{M}$
then all limit points of $\{\phi_p\}$ in a GEM are local maxima of L, and $L(\phi_p)$ converges monotonically to $L^* = L(\phi^*)$ for some local maxima ϕ^*

Convergence of an GEM sequence $\{\phi_p\}$

Convergence of an GEM sequence $\{\phi_p\}$ – 1

Define two sets:

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

$$\mathscr{M}(a) = \{\phi \in \mathscr{M} : L(\phi) \equiv a\}$$

If $L(\phi_p \to L^*)$, then the linit points of ϕ_p are in $\varphi(L^*)$ (or $\mathscr{M}(L^*)$)

Theorem 4

ϕ_p is a GEM sequence generated by $\phi_{p+1} \in M(\phi_p)$, and suppose:
1) M is a closed p2s map over φ^C (or $\mathscr{M}^C$)
2) $L(\phi_{p+1}) > L(\phi_p)\ \forall \phi_p \notin \varphi$ (or $\mathscr{M}$)
If $\varphi(L^*) = \{\phi^*\}$ (or $\mathscr{M}(L^*) = \{\phi^*\}$) where $L^* = \lim_{p\to\infty} L(\phi_p)$,
then $\lim_{p\to\infty} \phi_p = \phi^*$
NOTE: $\varphi(L^*) = \{\phi^*\}$ (or $\mathscr{M}(L^*) = \{\phi^*\}$) can be relaxed by $\lim_{p\to\infty} ||\phi_{p+1} - \phi_p|| = 0$ (see Theorem 5)

Convergence of an GEM sequence $\{\phi_p\}$ – 2

Theorem 5

ϕ_p is a GEM sequence generated by $\phi_{p+1} \in M(\phi_p)$, and suppose:
1) M is a closed p2s map over φ^C (or $\mathscr{M}^C$)
2) $L(\phi_{p+1}) > L(\phi_p)$ $\forall \phi_p \notin \varphi$ (or $\mathscr{M}$)
If $\lim_{p \to \infty} ||\phi_{p+1} - \phi_p|| = 0$,
then all limit points of $\{\phi_p\}$ are in a connected and compact subset of $\varphi(L^*)$ or $\mathscr{M}(L^*)$ where $L^* = \lim_{p \to \infty} L(\phi_p)$.
(Here, a connected subset cannot be represented as an union of two disjoint sets.)
In particular, if $\varphi(L^*)$ or $\mathscr{M}(L^*)$ is discrete,
then ϕ_p converges to some ϕ^* in $\varphi(L^*)$ or $\mathscr{M}(L^*)$.
Proof see Theorem 28.1 of Ostrowski (1996) [1] and Theorem 1. ■

Convergence of an GEM sequence $\{\phi_p\}$ – 3

We have: $\varphi(L) = \{\phi \in \varphi : L(\phi) \equiv L\}$ and $\mathscr{M}(L) = \{\phi \in \mathscr{M} : L(\phi) \equiv L\}$
Define a set: $\psi(L) = \{\phi \in \Omega : L(\phi) \equiv L\}$

Theorem 6

ϕ_p is a GEM sequence generated by $\phi_{p+1} \in M(\phi_p)$ with $\nabla Q(\phi_{p+1}|\phi_p) = 0$, and suppose $\nabla Q(\phi^{'}|\phi)$ is continuous in $\phi^{'}$ and ϕ.
If either (a) $\psi(L^*) = \{\phi^*\}$, or
(b) $\lim_{p\to\infty} ||\phi_{p+1} - \phi_p|| = 0$ and $\psi(L^*)$ is discrete satisfies,
then ϕ_p converges to a stationary point ϕ^* with $L(\phi^*) = L^*$, the limit of $L(\phi_p)$.

Proof

The continuity of $\nabla Q(\phi^{'}|\phi)$ implies condition 1) and 2) of Theorem 1.
(a) or (b) in Theorem 6 is stronger than the condition after if in Theorem 4 or Theorem 5.
The continuity of $\nabla Q(\phi^{'}|\phi)$ and $\nabla Q(\phi_{p+1}|\phi_p) = 0$ imply
$\nabla L(\phi^*|\phi^*) = \nabla Q(\phi^*|\phi^*) = 0.$ ■

Convergence of an GEM sequence $\{\phi_p\}$ – 4

Theorem 7

Suppose that $L(\phi)$ is unimodal in Ω with ϕ^* being the only stationary point and that $\nabla Q(\phi^{'}|\phi)$ is continuous in ϕ and $\phi^{'}$.
Then for any GEM sequence $\{\phi_p\}$, ϕ_p converges to the unique maximizer ϕ^* of $L(\phi)$.

NOTE: Theorem 7 is a corollary of Theorem 6.

Summary

Summary

- If the unobserved data can be expressed by a curved exponential family with compact parameter space, then all the limit points of a GEM sequence $\{\phi_p\}$ are stationary points of the likelihood function $L(\phi)$[3].
- If the likelihood $L(\phi)$ is unimodal with the only stationary point ϕ^*, and $\nabla Q(\phi^{'}|\phi)$ is continuous in $\phi^{'}$ and ϕ, then any GEM sequence $\{\phi_p\}$ converges to the unique maximizer ϕ^* of $L(\phi)$.

Questions?

References i

H. AS.
Ostrowski, am-solution of equations and systems of equations, 1967.

A. P. Dempster, N. M. Laird, and D. B. Rubin.
Maximum likelihood from incomplete data via the em algorithm.
Journal of the Royal Statistical Society: Series B (Methodological), 39(1):1–22, 1977.

C. J. Wu.
On the convergence properties of the em algorithm.
The Annals of statistics, pages 95–103, 1983.