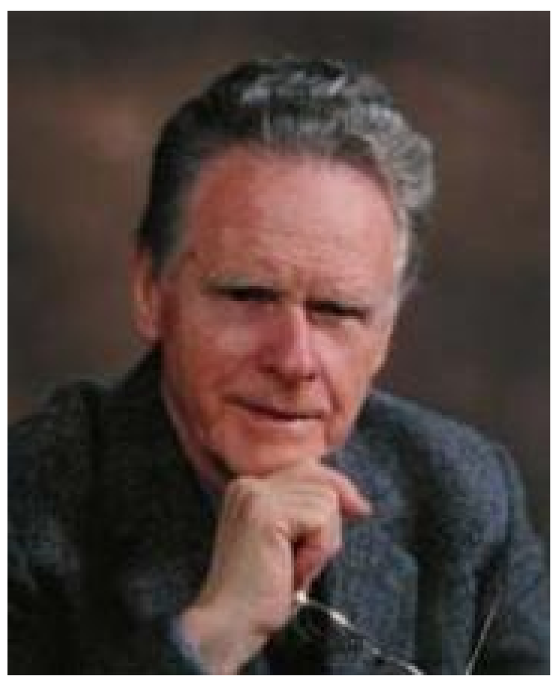
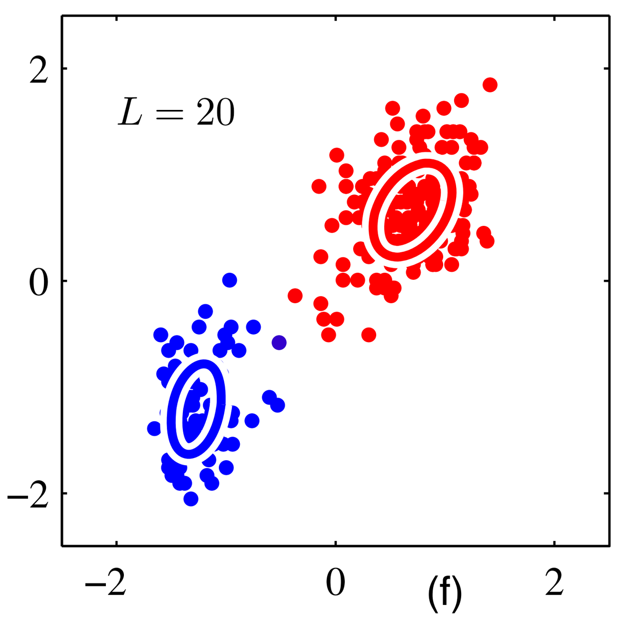
From Dempster to Lyapunov: A Dynamical Systems Approach to EM Algorithm Convergence

Orlando Romero **Sarthak Chatterjee**

Rensselaer Polytechnic Institute

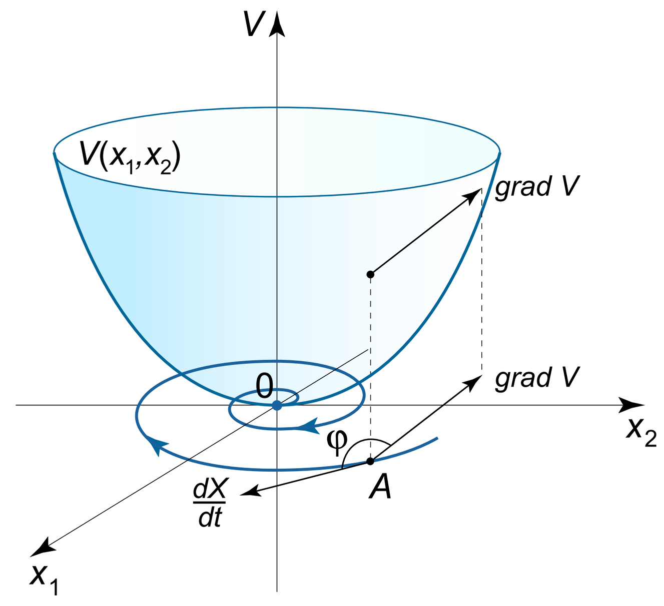


Arthur P. Dempster (1929 –)



EM Classification Example

Aleksandr Lyapunov (1857 – 1918)



Lyapunov Function Example

Framework – Parameter Estimation

- **Incomplete data:** $y \in \mathbb{R}^m$
- **Complete data:** $z = (x, y)$, $x \in \mathcal{X}$ is latent
- **Unknown parameter:** $\theta \in \Theta$
- **Statistical model:** $\{p_\theta(x, y) : \theta \in \Theta, x \in \mathcal{X}\}$
- $\hat{\theta}_{\text{ML}} \overset{\text{def}}{=} \underset{\theta\in\Theta}{\text{argmax}}\ \mathcal{L}(\theta)$, where $\mathcal{L}(\theta) = p_\theta(y) = \int_{\mathcal{X}} p_\theta(x, y)\,\mathrm{d}x$
- How to compute $\hat{\theta}_{\text{ML}}$?

Expectation-Maximization (EM) Algorithm

- $Q(\theta, \theta') \overset{\text{def}}{=} \mathbb{E}_{p_{\theta'}(x|y)}[\log p_\theta(x, y)] = \displaystyle\int_{\mathcal{X}} \frac{p_{\theta'}(x, y)}{p_{\theta'}(y)} \log p_\theta(x, y)\,\mathrm{d}x.$
- 1: **initialize** $\theta_0 \in \Theta$
  2: **for** $k = 1, 2, \ldots$ **do**
  3: **E-step:** Compute $Q(\theta, \theta_k)$
  4: **M-step:** Determine $\theta_{k+1} = \underset{\theta\in\Theta}{\text{argmax}}\, Q(\theta, \theta_k)$
  5: **end for**
  6: **return** θ_∞
- **Assumptions:**
  - $p_\theta(y) > 0$ for every $\theta \in \Theta$.
  - $\mathcal{X} = \{x \in \mathbb{R}^n : p_\theta(x, y) > 0\}$ does not depend on $\theta \in \Theta$.
  - for each $\theta' \in \Theta$, the function $Q(\cdot, \theta')$ as a unique global maximizer.
  - $\mathcal{L}(\theta)$ is twice continuously differentiable.
  - $\theta \mapsto p(\cdot|y)$ is injective.

Dynamical Systems Overview

- State-space model: $\theta[k+1] = F(\theta[k])$ with $\theta[0] = \theta_0$.
- $\theta^\star$ is an *equilibrium* if $\theta_0 = \theta^\star \implies \theta[k] = \theta^\star$ for every k.
- $\theta^\star$ is a *stable* equilibrium if $\forall\varepsilon > 0, \exists\delta > 0$ such that $\|\theta_0 - \theta^\star\| \le \delta \implies \|\theta[k] - \theta^\star\| \le \varepsilon$ for every k.
- $\theta^\star$ is an *asymptotically stable* equilibrium if it is stable and $\exists\delta > 0$ such that $\|\theta_0 - \theta^\star\| \le \delta \implies \lim_{k\to\infty} \theta[k] = \theta^\star$.
- $\theta^\star$ is an *exponentially stable* equilibrium if it is stable and $\exists\delta, c, \gamma > 0$ such that $\|\theta_0 - \theta^\star\| \le \delta \implies \|\theta[k] - \theta^\star\| \le c \cdot e^{-\gamma k}\|\theta_0 - \theta^\star\|$ for every k.

A Dynamical Systems Interpretation of EM

- $F(\theta') = F^{\text{EM}}(\theta') = \text{argmax}_{\theta\in\Theta}\, Q(\theta, \theta') \implies \theta[k] = \theta_k$.
- $\theta^\star$ is an equilibrium for $F = F^{\text{EM}} \iff \theta^\star$ is a fixed point of EM .
- $\theta^\star$ is asymptotically stable $\implies$ EM is locally convergent to $\theta^\star$.

Lyapunov Theorem

Let $\theta^\star$ be an equilibrium. If there exists a continuous function $\mathcal{V} : \Theta \to \mathbb{R}$ (*Lyapunov function*) such that

- $\mathcal{V}$ is positive definite (w.r.t. $\theta^\star$), *i.e.* $\mathcal{V}(\theta^\star) = 0$ and $\mathcal{V}(\theta) > 0$ for $\theta \neq \theta^\star$;
- $-\Delta\mathcal{V}$ is positive definite, where $\Delta\mathcal{V}(\theta) \overset{\text{def}}{=} \mathcal{V}(F(\theta)) - \mathcal{V}(\theta)$,

then $\theta^\star$ is asymptotically stable. Furthermore, if $\mathcal{V}(\theta) \le a\|\theta - \theta^\star\|^2$ and $-\Delta\mathcal{V}(\theta) \ge b\|\theta - \theta^\star\|^2$, then $\theta^\star$ is exponentially stable, with $c = d/a$, $d = \lim_{\delta\to 0} \max_{\delta \ge \|\theta-\theta_0\| \ge \frac{1}{\delta}} \frac{\mathcal{V}(\theta)}{\|\theta-\theta^\star\|}$, and $\gamma = \log a - \log(a - b)$.

Main Results

- **Theorem 1:** If $\nabla^2\mathcal{L}(\hat{\theta}_{\text{ML}}) \prec 0$ then $\hat{\theta}_{\text{ML}}$ is asymptotically stable, and thus EM is locally convergent to $\hat{\theta}_{\text{ML}}$.
- **Proof:**
  - $Q(\theta, \theta') = \log\mathcal{L}(\theta) - \mathcal{D}_{\text{KL}}[p_{\theta'}(\cdot|y)\|p_\theta(\cdot|y)] - \mathcal{H}[p_{\theta'}(\cdot|y)]$
  - $F^{\text{EM}}(\hat{\theta}_{\text{ML}}) = \text{argmax}_{\theta\in\Theta}\{\log\mathcal{L}(\theta) - \mathcal{D}_{\text{KL}}[p_{\hat{\theta}_{\text{ML}}}(\cdot|y)\|p_\theta(\cdot|y)]\}$
  - ($\log\mathcal{L}(\theta), \mathcal{D}_{\text{KL}}[p_{\hat{\theta}_{\text{ML}}}(\cdot|y)\|p_\theta(\cdot|y)]$ are maximized at $\theta = \hat{\theta}_{\text{ML}}$) $\implies F^{\text{EM}}(\hat{\theta}_{\text{ML}}) = \hat{\theta}_{\text{ML}}$.
  - $\mathcal{V}(\theta) = \mathcal{L}(\hat{\theta}_{\text{ML}}) - \mathcal{L}(\theta)$ is positive definite w.r.t. $\hat{\theta}_{\text{ML}}$.
  - $\log\mathcal{L}(F^{\text{EM}}(\theta)) - \underbrace{\mathcal{D}_{\text{KL}}[p_\theta(\cdot|y)\|p_{F^{\text{EM}}(\theta)}(\cdot|y)]}_{>0\ (\theta\neq\hat{\theta}_{\text{ML}})} \ge \log\mathcal{L}(\theta) - \underbrace{\mathcal{D}_{\text{KL}}[p_\theta(\cdot|y)\|p_\theta(\cdot|y)]}_{=0} \implies -\Delta\mathcal{V}(\theta) = \mathcal{L}(F^{\text{EM}}(\theta)) - \mathcal{L}(\theta)$ is also positive definite w.r.t. $\hat{\theta}_{\text{ML}}$.
- **Theorem 2**: If $\theta^\star$ is a limit point of EM such that $\nabla^2\mathcal{L}(\theta^\star) \prec 0$, then it is asymptotically stable and thus EM is locally convergent to $\theta^\star$.
- **Proof**:
  - (F^{EM} is continuous and $\theta^\star$ is a limit point) $\implies F(\theta^\star) = \theta^\star$.
  - Repeat same argument for $\mathcal{V}(\theta) = \mathcal{L}(\theta^\star) - \mathcal{L}(\theta)$ w.r.t. $\theta^\star$ in a small enough open ball around $\theta^\star$.
- **Theorem 3:** In the same conditions of Theorem 2, and assuming $\mathcal{L}(F^{\text{EM}}(\theta)) - \mathcal{L}(\theta) \ge b\|\theta - \theta^\star\|^2$ for some $b > 0$, then $\theta^\star$ is exponentially stable and thus the linear convergence rate of EM can be explicitly bounded in terms of $\mathcal{L}(\theta)$.

O. Romero, S. Chatterjee, and S. Pequito, "Convergence of the Expectation-Maximization Algorithm Through Discrete-Time Lyapunov Stability Theory" (under review).