한국통계학회논문집 제3권 2호
1996년 8월 pp. 239-247

PEM 그래디언트 알고리즘에 관한 연구

김 승 구[1)]

요 약

본 연구에서는 페널티化 대수우도함수의 해를 구하기 위해 Lange (1995)의 EMG알고리즘을 적용할 경우에 발생하는 문제점을 제시하고 이를 해결하기 위해 OSLG라 명명된 알고리즘을 제안한다. OSLG알고리즘은 EMG알고리즘이나 Green (1990)의 OSL알고리즘으로 해결할 수 없는 문제에 쉽게 적용된다. 한편 이 알고리즘은 EMG알고리즘의 변형이지만 OSL알고리즘과 같은 국소수렴성질을 갖는다. OSLG알고리즘은 특히 페널티함수에 대한 2차 도함수행렬이 대각행렬이 아닌 응용분야에서 매우 유용하게 사용될 수 있을 것으로 기대된다.

1. 서론 및 기본개념

Dempster (1977) 로부터 소개되기 시작한 EM알고리즘은 현대통계학에서 최우추정의 일반적 해법으로 인식될 정도로 매우 자주 사용되고 있다. 실제로 최우추정치를 구하는데 있어서 「뉴튼」알고리즘과 같은 기존의 해법으로는 해결이 불가능해 보이는 문제가 EM알고리즘으로 쉽게 해결되는 경우를 Little (1987) 등에서 많이 찾아 볼 수 있다. 먼저 EM알고리즘의 개념을 간략히 살펴보자.

y를 관측가능한 불완전자료 그리고 x를 관측불능의 완전자료라 할때 $y=y(x)$와 같은 통계량으로 나타낸다. 그리고 y와 x의 확률밀도를 각각 $g(y|\theta)$ 및 $f(x|\theta)$라 하고 모수벡터 θ에 대한 대수우도를 $L(\theta)=\ln\{g(y|\theta)\}$를 최대화하는 MLE를 $\hat{\theta}$로 나타내자.

이때 EM알고리즘은 $n+1$ 단계에서, 주어진 θ^n에 대해 정규방정식

$$D^{10}Q(\theta|\theta^n)=0 \tag{1.1}$$

을 만족하는 $\theta=\theta^{n+1}$를 찾는 반복알고리즘이다. 여기서 $Q(\theta|\theta^n)=E(\ln\{f(x|\theta)\}|y,\theta^n)$ 이고 $D^{ij}Q(x|y)=\partial^{i+j}Q(x|y)/\partial^i x\partial^j y$를 나타내는 미분연산자이다. 흔히 이해를 돕기위해 E-단계와 M-단계로 구분하여 나타내나, 본 논문에서는 두 단계를 결합해서 표현하기로 하겠다.

그런데 만약 방정식 (1.1)으로 부터 θ에 대한 전개식을 쉽게 얻을 수 없다면 EM알고리즘의 사용은 사실상 불가능하게 된다. 이때 이 문제를 해결하는 가장 설득력 있는 방법은 Lange (1995)의

1) (220-702) 강원도 원주시 우산동 산41 상지대학교 통계학과, 조교수.

EMG알고리즘(EM gradient algorithm)

$$\theta^{n+1}=\theta^{n}-\{D^{20}Q(\theta^{n}|\theta^{n})\}^{-1}D^{10}Q(\theta^{n}|\theta^{n}) \tag{1.2}$$

을 사용하는 것이다. 이 방법은 EM알고리즘의 수렴속도가 매우 느리다(거의 선형적)는 점에 착안하여, 각 M-단계에서 「뉴튼」알고리즘의 1-단계 추정치를 사용한 알고리즘으로서 EM알고리즘과 같은 수렴성질을 갖는다. 또 식(1.2)로부터 알 수 있듯이 $\theta=\theta^{n+1}$에 대한 전개가 필요치 않다. 그리고 대부분의 문제에서 $D^{20}Q(\theta^{n}|\theta^{n})$은 역행렬 계산의 노력이 필요없는 대각행렬이라는데 이 알고리즘의 유용성이 있다. 본 논문에서도 앞으로 이 행렬을 대각행렬로 가정할 것이다.

한편 페널티化 대수우도(penalized log-likelihood)

$$L(\theta)-\lambda J(\theta) \tag{1.3}$$

을 최대화하는 추정량 $\hat{\theta}$을 MPL(maximum penalized likelihood)추정량 혹은 θ에 대한 사전밀도가 $p(\theta)\propto\exp\{-\lambda J(\theta)\}$인 MAP(maximum *a posteriori*)추정량이라 한다. 여기서 λ는 양정수로서 흔히 페널티함수 $J(\theta)$에 대한 평활상수라 부른다.

이때 PEM알고리즘(penalized EM algorithm)이란 주어진 θ^{n}에 대해

$$D^{10}Q(\theta|\theta^{n})-\lambda DJ(\theta)=0 \tag{1.4}$$

을 만족하는 $\theta=\theta^{n+1}$을 반복적으로 결정하는 알고리즘이다. 그리고 식 (1.4)에 대응하는 EMG알고리즘은

$$\theta^{n+1}=\theta^{n}-\{D^{20}Q(\theta^{n}|\theta^{n})-\lambda D^{2}J(\theta^{n})\}^{-1}\{D^{10}Q(\theta^{n}|\theta^{n})-\lambda DJ(\theta^{n})\} \tag{1.5}$$

과 같이 얻을 수 있다. 이 알고리즘을 PEMG알고리즘(penalized EM gradient algorithm)이라 부르기로 하자.

많은 통계학 문제에서 PEMG알고리즘은 정규방정식 (1.4)의 대안으로 사용될 수 있을 것이다. 그러나 Abdalla(1990)등에서처럼 공간종속적 자료를 다루는 화상복원과 같은 응용분야에서는 $J(\theta)$가 θ의 공간정보를 설명하는 불균일도 페널티로 주어지므로 $D^{2}J(\theta^{n})$는 비대각행렬이 된다. 더우기 이 행렬의 차수는 매우 크기 때문에 계산시간의 장애로 인해 대각화변환 등의 방법은 실효성이 없다. 결국 $D^{20}Q(\theta^{n}|\theta^{n})$이 대각행렬이라 하더라도 $D^{2}J(\theta^{n})$이 비대각행렬이면 PEMG알고리즘은 EMG알고리즘에서와 같은 유용성을 기대할 수 없게 된다.

이러한 문제점을 해결하기 위해 본 연구에서는 PEMG알고리즘을 변형한 다음의 알고리즘을 제안한다.

$$\theta^{n+1}=\theta^{n}-\{D^{20}Q(\theta^{n}|\theta^{n})\}^{-1}\{D^{10}Q(\theta^{n}|\theta^{n})-\lambda DJ(\theta^{n})\}\ . \tag{1.6}$$

이 알고리즘은 단순히 비대각행렬인 $\lambda D^{2}J(\theta^{n})$을 PEMG알고리즘의 향행렬로부터 제거한 식이다. 결국 향행렬이 대각행렬이 되어 역행렬계산을 피할 수 있게 해준다. 우리는 이 알고리즘을

OSLG알고리즘(OSL gradient algorithm)이라 부르도록 하겠다. 그 이유는 이 알고리즘의 국소수렴성이 Green (1990)의 OSL알고리즘(one-step lated algorithm)의 그 것과 정확히 일치하기 때문이다.

본 연구의 2절에서 OSLG알고리즘과 OSL알고리즘의 관계를 살펴보고 3절에서는 PEMG알고리즘과 OSLG알고리즘 수렴성질을 밝힌다. 그리고 4절에서는 적용사례를 예시하며, 마지막으로 5절에서는 결론과 아울러 제안된 알고리즘이 보완되어야 할 점에 대해 추가적으로 논의한다.

2. OSL알고리즘과 OSLG알고리즘의 관계

OSL알고리즘은 방정식 $D^{10}Q(\theta|\theta^n)-\lambda DJ(\theta)=0$ 대신

$$D^{10}Q(\theta|\theta^n)-\lambda DJ(\theta^n)=0 \tag{2.1}$$

을 만족하는 반복해를 구하는 알고리즘이다. 이 것은 도함수 $DJ(\theta)$가 θ에 대해 복잡한 형태를 가질때 $DJ(\theta^n)$으로 대체하므로써 θ에 대한 전개가 쉽게 이루어지도록 만들어진 알고리즘이다.

그런데, 만약 $D^{10}Q(\theta|\theta^n)$도 역시 θ에 대한 전개의 어려움이 따른다면 다음과 같은 방법을 고려할 수 있을 것이다. 즉, $\theta=\theta^n$ 에서의 $D^{10}Q(\theta|\theta^n)$에 대한 근사식

$$D^{10}Q(\theta|\theta^n) \approx D^{10}Q(\theta^n|\theta^n)+D^{20}Q(\theta|\theta^n)(\theta-\theta^n)$$

을 식 (2.1)에 대입하여 얻은 방정식

$$D^{20}Q(\theta|\theta^n)(\theta-\theta^n)+D^{10}Q(\theta^n|\theta^n)-\lambda DJ(\theta^n) \approx 0 \tag{2.2}$$

로 부터 θ에 대한 근사해를 구하는 것이다. 이 것은 곧 식 (1.6)의 OSLG알고리즘을 의미한다. 이 사실로 부터, OSLG알고리즘은 OSL알고리즘을 그래디언트형식화한 알고리즘이라 할 수 있다. 또한 OSL와 EMG알고리즘의 이점을 공유하고 있는, 다시 말해서, 함수 $D^{10}Q(\theta|\theta^n)$와 $DJ(\theta)$에서 발생하는 전개상의 난점을 동시에 해결하는 알고리즘이라 하겠다.

3. OSLG 및 PEMG알고리즘의 수렴성질

불완전관측자료 y에 대한 완전자료 x의 조건부확률밀도를 $k(x|y,\theta)$라 하자. 이때 $y=y(x)$이면, MAP추정량 $\hat{\theta}$은 사후밀도

$$p_o(\theta|x)\propto f(x|\theta)p(\theta)=g(y|\theta)p(\theta)k(x|y,\theta) \tag{3.1}$$

을 최대로 한다. 식 (3.1)에서 등식의 양변에 대수를 취한 후 조건부기대값을 취하면,

$$Q(\theta|\theta^n) - \lambda J(\theta) = \{L(\theta) - \lambda J(\theta)\} + H(\theta|\theta^n) \tag{3.2}$$

와 같이 나타낼 수 있다. 단, $H(\theta|\theta^n) = E(\ln\{k(x|y,\theta)\}|y,\theta^n)$.
여기서 $Q_p(\theta) = Q(\theta|\theta^n) - \lambda J(\theta)$ 및 $L_p(\theta) = L(\theta) - \lambda J(\theta)$라 놓자. 이때 잘 알려진 관계식

$$H(\theta|\theta^n) \leq H(\theta^n|\theta^n) \tag{3.3}$$

으로 부터 $Q_p(\theta|\theta^n)$을 최대로 하는 $\theta = \theta^{n+1}$에 대해

$$L_p(\theta^{n+1}) > L_p(\theta^n) \tag{3.4}$$

이 항상 성립함을 보일 수 있다. 이는 PEM알고리즘의 단계가 진행됨에 따라 페널티화 대수우도를 항상 증가시키게 되어, 만약 θ^n이 $\hat{\theta}$의 근방에 놓이게 되면 결국 θ^∞가 $\hat{\theta}$로 수렴하게 됨을 의미한다. 이로부터 OSLG알고리즘과 PEMG알고리즘이 식 (3.4)의 관계를 만족하는지를 살펴보기로 하겠다.

먼저 $D^2L(\theta^n)$과 $D^{20}H(\theta^n|\theta^n)$을 각각 음정치행렬 및 음반정치행렬이라 하면, 식 (3.2)의 관계로부터 $D^{20}Q(\theta^n|\theta^n)$는 음정치행렬임을 알 수 있다. 그리고 $D^2J(\theta^n)$는 양반정치행렬이라 하고 모두는 같은 차수의 대칭행렬이라 하자.

정리 1. 행렬 $-D^{20}H(\theta^n|\theta^n) - \lambda D^2J(\theta^n)$이 최소한 양반정치행렬이 되도록 하는 모든 λ에 대해, OSLG알고리즘은 $L_p(\theta^{n+1}) > L_p(\theta^n)$를 만족한다.

(증명) $\theta = \theta^n$에서 $L_p(\theta^{n+1})$에 대한 Taylor 2차 전개식

$$\begin{aligned} L_p(\theta^{n+1}) &= L_p(\theta^n) + DL_p(\theta^n)^T(\theta^{n+1} - \theta^n) \\ &\quad + (1/2)(\theta^{n+1} - \theta^n)^T D^2L_p(\theta^n)(\theta^{n+1} - \theta^n) + o((\theta^{n+1} - \theta^n)^T(\theta^{n+1} - \theta^n)) \end{aligned}$$

으로부터, $\Delta L_p(\theta^n) \equiv L_p(\theta^{n+1}) - L_p(\theta^n)$라 하면

$$\begin{aligned} \Delta L_p(\theta^n) = &\ DL_p(\theta^n)^T(\theta^{n+1} - \theta^n) + (1/2)(\theta^{n+1} - \theta^n)^T D^2L_p(\theta^n)(\theta^{n+1} - \theta^n) \\ &+ o((\theta^{n+1} - \theta^n)^T(\theta^{n+1} - \theta^n)). \end{aligned} \tag{3.5}$$

그런데 식 (3.3)은 $\theta = \theta^n$ 일때 $H(\theta|\theta^n)$이 최대임을 말하고있다. 따라서 식 (3.2)에서 $D^{10}Q(\theta^n|\theta^n) = DL(\theta^n)$이 성립한다. 결국 식 (1.6)의 OSLG알고리즘은

$$\theta^{n+1}-\theta^{n}= -\{D^{20}Q(\theta^{n}|\theta^{n})\}^{-1}\{D^{10}Q(\theta^{n}|\theta^{n})-\lambda DJ(\theta^{n})\}$$
$$= -\{D^{20}Q(\theta^{n}|\theta^{n})\}^{-1}DL_{p}(\theta^{n})$$

으로 나타낼 수 있다. 여기서 $r=D^{20}Q(\theta^{n}|\theta^{n})^{-1}DL_{p}(\theta^{n})$라 하고 이 결과를 식 (3.5)에 대입하면

$$\begin{aligned}\Delta L_{p}(\theta^{n}) &= -DL_{p}(\theta^{n})^{T}r + (1/2)r^{T}D^{2}L_{p}(\theta^{n})r + o(r^{T}r)\\ &= (1/2)r^{T}\{D^{2}L_{p}(\theta^{n})-2D^{20}Q(\theta^{n}|\theta^{n})\}r + o(r^{T}r)\\ &= (1/2)r^{T}Kr + o(r^{T}r)\end{aligned}$$

이 된다. 단, $K=D^{2}L_{p}(\theta^{n})-2D^{20}Q(\theta^{n}|\theta^{n})$ 이다.

그런데 식 (3.2)의 관계로 부터, $K=-D^{20}Q(\theta^{n}|\theta^{n})+\{-D^{20}H(\theta^{n}|\theta^{n})-\lambda D^{2}J(\theta^{n})\}$임을 알 수 있다. 따라서 $-D^{20}H(\theta^{n}|\theta^{n})-\lambda D^{2}J(\theta^{n})$가 음정치행렬이 될정도로 λ가 결정되지만 않는다면 임의의 벡터 x에 대해 $x^{T}Kx \geq -x^{T}D^{20}Q(\theta^{n}|\theta^{n})x > 0$ 을 만족하며 $|o(x^{T}x)| \leq (1/2)x^{T}Kx$ 이므로 결국 $\Delta L_{p}(\theta^{n})>0$ 이다. □

사실상 정리 1.의 조건은 Green (1990)에서 제시된 OSL알고리즘의 수렴조건이다. 또한 통계학 대부분의 문제에서 $-D^{20}H(\theta^{n}|\theta^{n})-\lambda D^{2}J(\theta^{n})$가 음정치행렬이 되도록 λ가 주어지는 경우는 극히 드물다. 결국 정리 1은 일반성을 크게 잃지 않고 거의 모든 통계학 문제에서 OSLG알고리즘이 $L_{p}(\theta^{n+1})>L_{p}(\theta^{n})$이 만족됨을 보장한다 할 수 있다. 더우기, 혹시 모든 λ에 대해 조건행렬이 음정치인 경우라도 θ에 대한 재모수화를 통해 양반정치화 할 수도 있을 것이다.

정리 2. 모든 λ에 대해, PEMG알고리즘은 $L_{p}(\theta^{n+1})>L_{p}(\theta^{n})$를 만족한다.

(증명) 정리 1.과 비슷한 방법으로 전개하면,

$$\theta^{n+1}-\theta^{n}=\{D^{20}Q_{p}(\theta^{n}|\theta^{n})\}^{-1}DL_{p}(\theta^{n})$$

을 얻는다. 이 것을 식(3.5)에 대입하면 $\Delta L_{p}(\theta^{n})=r^{T}\{D^{2}L_{p}(\theta^{n})-2D^{20}Q_{p}(\theta^{n}|\theta^{n})\}r$ 이 된다. 단, $r=D^{20}Q_{p}(\theta^{n}|\theta^{n})^{-1}DL_{p}(\theta^{n})$. 그런데 식 (3.2)의 관계로부터

$$D^{2}L_{p}(\theta^{n})-2D^{20}Q_{p}(\theta^{n}|\theta^{n})=-D^{20}H(\theta^{n}|\theta^{n})-D^{20}Q(\theta^{n}|\theta^{n})+\lambda D^{2}J(\theta^{n})$$

이다. 다시말해서 우변은 모든 $\lambda>0$에 대해 양정치행렬이다. 결국 $\Delta L_{p}(\theta^{n})>0$ 이다. □

두 정리로부터, OSLG 및 PEMG알고리즘은 M-단계가 진행됨에 따라 페널티화 대수우도를 증가시킨다는 사실을 알 수 있었다. 이는 곧 두 알고리즘은 만약 θ^n이 $\tilde{\theta}$의 근방에 놓이게 되면 같은 고정점 $\theta^\infty = \tilde{\theta}$에 수렴함을 의미한다.

다음은 두 알고리즘에 대한 $\tilde{\theta}$ 근방에서의 수렴률을 살펴보도록 한다. 여기서 OSLG알고리즘을 $\theta^{n+1} = M(\theta^n)$이라 하고 PEMG알고리즘은 $\theta^{n+1} = N(\theta^n)$이라 놓자.

먼저 OSLG알고리즘의 수렴률을 살펴보기 위해 $M(\theta^n)$을 θ^n에 대해 미분한 후 $\theta^n = \theta^{n+1} = \tilde{\theta}$을 대입하면, $DL_p(\tilde{\theta}) = 0$이란 사실로부터

$$\begin{aligned} DM(\tilde{\theta}) &= I - D^{20}Q(\tilde{\theta}|\tilde{\theta})^{-1}D^2L_p(\tilde{\theta}) + D\{D^{20}Q(\tilde{\theta}|\tilde{\theta})^{-1}\}DL_p(\tilde{\theta}) \\ &= -D^{20}Q(\tilde{\theta}|\tilde{\theta})^{-1}\{-D^{20}Q(\tilde{\theta}|\tilde{\theta}) + D^2L(\tilde{\theta}) - \lambda D^2J(\tilde{\theta})\} \\ &= -D^{20}Q(\tilde{\theta}|\tilde{\theta})^{-1}\{-D^{20}H(\tilde{\theta}|\tilde{\theta}) - \lambda D^2J(\tilde{\theta})\} \end{aligned} \tag{3.6}$$

을 구할 수 있다. 한편, 비슷한 방법으로 PEMG알고리즘의 수렴률은

$$DN(\tilde{\theta}) = \{-D^{20}Q(\tilde{\theta}|\tilde{\theta})^{-1} + \lambda D^2J(\tilde{\theta})\}\{-D^{20}H(\tilde{\theta}|\tilde{\theta})\} \tag{3.7}$$

과 같이 얻을 수 있다.

Green (1990)은 $\tilde{\theta}$의 근방에서 OSL알고리즘이 PEM알고리즘보다 더 빠르게 수렴함을 보였다. 그런데 $DM(\tilde{\theta})$와 $DN(\tilde{\theta})$은 각각 OSL알고리즘 및 PEM알고리즘의 수렴률과 정확히 일치한다는 것이다. 이는 곧 OSLG알고리즘이 $\tilde{\theta}$의 근방에서 PEMG알고리즘보다 오히려 더 빠르게 수렴함을 의미한다.

4. 응용예제

<u>포아송 광방출 토머그래피</u>(Poisson photon emission tomography)

의료영상촬영을 위해 이용되는 방광출 토머그래피에 EM알고리즘이 어떻게 적용되는지 간단히 살펴 보도록 하자. 먼저 방사성의약품이 환자의 환부에 투약된다. 이 환부로부터 입자가 방출될 때 i번째 단위측정영역(bins)에서 관측된 입자수를 y_i라 하자. 그리고 위치 i에서의 관측에 영향을 미치는 영역을 N_i라 하면,

$$y_i \sim Poisson(\sum_{j \in N_i} b_{ij}\theta_j)$$

인 것으로 알려져 있다. 여기서 $m \times p$ 행렬 $B = \{b_{ij}\}$는 기지의 점확산행렬(point-spread matrix)로서 기기의 특성으로 인한 영상의 퍼짐상태(blurring)를 설명한다.

이때, 완전자료 x_{ij}는 환부의 j번째 위치에서 방출된 광자들을 i번째 단위측정영역에서 개별적으로 측정한 입자수로 정의한다. 그러나 x_{ij}의 측정은 현실적으로 불가능하다.

결국

$$y_i = \sum_{j \in N_i} x_{ij} = \sum_{j=1}^{p} x_{ij}\ ,\quad i=1,\cdots,m$$

로 나타낼 수 있으며, $x_{ij} \sim Poisson(b_{ij}\theta_j)$라 하면 y_i의 확률모형은 보존된다. 여기서 완전자료 $x=\{x_{ij}\}$와 불완전자료 $y=\{y_i\}$에 대해 $\theta=\{\theta_{ij}\}$는 결측자료로 정의된다.

한편, 일반적으로 모수벡터 θ에 대한 사전정보는

$$p(\theta) \propto \exp\{-\lambda J(\theta)\}$$

와 같은 MRF(markov random field)모형으로 주어진다. 여기서 $J(\theta)=\sum_{j\sim k} V(\theta_j, \theta_k)$를 흔히 에너지함수라 부르는데, $j\sim k$는 i에 대한 쌍별근린(pairwise-nearest neighbors)을 의미하고, $V(\theta_j, \theta_k)$는 $\theta_j - \theta_k$에 대한 2차함수, 절대값 혹은 복잡한 초월함수 등으로 주어진다.

여기서 우리의 목적은 사후밀도를 최대화하는 θ를 구하는 것이다. 이 경우 직접 불완전자료에 의한 정규방정식 $DL(\theta) - \lambda DJ(\theta) = 0$으로부터 해를 유도할 수 없다.

또한 m, p는 매우 큰 값이기 때문에 비대각행렬인 $D^{20}Q(\theta^n|\theta^n) - \lambda D^2 J(\theta^n)$의 역행렬을 직접 계산한다는 것은 거의 불가능하다. 따라서 이 문제에 PEMG알고리즘을 적용할 수는 없다. 그러나 OSLG알고리즘 사용하면 문제가 쉽게 해결된다. 왜냐하면 비대각성분인 $\lambda D^2\{\sum_k V(\theta_j^n, \theta_k^n)\}$의 역행렬을 계산할 필요가 없기 때문이다.

먼저 E-단계에서 $E(x_{ij}|y_i, \theta^n)$의 추정치는

$$\hat{x}_{ij} = y_i b_{ij}\theta_j^n \Big/ \sum_{k=1}^{p} b_{ik}\theta_k^n \tag{4.1}$$

이고 M-단계에서

$$\begin{aligned}\{D^{10}Q(\theta^n|\theta^n)\}_j &= \frac{\partial}{\partial\theta_j} E(\ln\{f(x|\theta)\}|y, \theta^n)\ |_{\theta_j=\theta_j^n} \\ &= \frac{\partial}{\partial\theta_j} \sum_i^m \sum_j^p \{\ \hat{x}_{ij}\ln\{-b_{ij}\theta_j\} - b_{ij}\theta_j\}\ |_{\theta_j=\theta_j^n} \\ &= \sum_i^m \hat{x}_{ij}/\theta_j^n - \sum_i^m b_{ij}\end{aligned}$$

과, 한 번더 미분하여 $p\times p$ 대각행렬 $D^{20}Q(\theta^n|\theta^n) = diag\left\{-\sum_i^n \hat{x}_{ij} \Big/ (\theta_j^n)^2\right\}$을 얻을 수 있다.

따라서 θ^{n+1}의 j번째 성분은

$$\begin{aligned}\theta_j^{n+1} &= \theta_j^n - \{D^{20}Q(\theta^n|\theta^n)\}_{jj}^{-1}\{D^{10}Q(\theta^n|\theta^n) - \lambda DJ(\theta^n)\}_j \\ &= \theta_j^n + \frac{(\theta_j^n)^2}{\sum_i^m \hat{x}_{ij}}\left\{\frac{\sum_i^m \hat{x}_{ij}}{\theta_j^n} - \sum_i^m b_{ij} - \lambda\{DJ(\theta^n)\}_j\right\} \\ &= \theta_j^n\left\{2 - \left(\sum_i^m b_{ij} + \lambda\{DJ(\theta^n)\}_j\right) \Big/ \sum_i^m\left(y_i b_{ij} / \sum_k^p b_{ik}\theta_k^n\right)\right\} \\ &\quad , \ j=1,\cdots,p\end{aligned} \tag{4.2}$$

과 같이 결정된다.

5. 결론 및 추가논의

본 연구에서는 페널티함수의 2차도함수행렬 $D^2J(\theta^n)$에 기인한 역행렬계산상의 어려움 때문에 페널티화 대수우도에 대해 PEMG알고리즘의 적용이 불가능 할 때, OSLG알고리즘을 이용하면 문제가 쉽게 해결될 수 있음을 소개하였다. 제안된 OSLG알고리즘은 Lange (1995)의 EMG알고리즘을 변형한 것으로서 $\theta^\infty = \hat{\theta}$에서의 국소수렴성질이 Green (1990)의 OSL알고리즘과 일치하며, PEMG알고리즘보다 빠르게 수렴함을 보였다. 특히 이 알고리즘은 화상분석과 같은 응용분야에서 다루기 쉽지 않은 수식으로 구성된 EM알고리즘을 해결하는데 유용하게 사용될 것으로 기대된다.

식 (3.6)에서 알 수 있듯이 결측정보(missing information) $-D^{20}H(\hat{\theta}|\hat{\theta})$가 상대적으로 큰 응용분야에서 OSLG알고리즘은 EM알고리즘과 마찬가지로 매우 느리게 수렴한다. 따라서 수렴속도를 가속시키는 개량이 요구된다.

여기서 현재 우리가 연구하고 있는 2가지 방법을 소개한다. 우선 OSLG알고리즘의 수렴속도를 좀 더 향상시키기 위하여, Lange (1995)에서 처럼

$$\theta^{n+1} = \theta^n - \gamma^n\{D^{20}Q(\theta^n|\theta^n)\}^{-1}\{D^{10}Q(\theta^n|\theta^n) - \lambda DJ(\theta^n)\} \tag{5.1}$$

과 같이 가변보폭(variable-step length) γ^n의 삽입하는 것이다. 참고로, Lange (1995)는 자료의 결측비율이 클 때 식 (1.2)에 $\gamma^n \approx 2$의 사용을 제안하였는데, 식 (5.1)에 대해서도 수렴속도를 최적으로 하는 γ^n을 결정하기 위한 연구가 요구된다. OSLG알고리즘을 가속시키는 또 다른 방법으로 Louis의 터보알고리즘(Tanner, 1991)을 적용하는 것이다.

즉,

(1) M-단계:

$$\theta_{EM}^{n+1} = \theta^n - \{D^{20}Q(\theta^n|\theta^n)\}^{-1}\{D^{10}Q(\theta^n|\theta^n) - \lambda DJ(\theta^n)\}$$

(5.2)

(2) 가속 단계:

$$\theta^{n+1} = \theta^n + \{D^2L(\theta^n)\}^{-1}D^{20}Q(\theta^n|\theta^n)(\theta_{EM}^{n+1} - \theta^n)$$

과 같이 M-단계에 가속단계를 추가하여 (1), (2)식을 반복하는 방법이다.

참 고 문 헌

[1] Abdalla, M. and Kay, J. (1990). Edge Preserving Image Restoration, *Stochastic Models, Statistical Methods, and Algorithm in Image Analysis*, P. Barone *et al.*(Eds.), Springer-Verag, 1-13.

[2] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm(with discussion), *Journal of Royal Statistical Society* B, Vol. 39, 1-38.

[3] Green, P.J. (1990). On use of the EM algorithm for penalized likelihood estimation, *Journal of Royal Statistical Society* B, Vol. 52, 443-452.

[4] Lange, K. (1995). A Gradient Algorithm Locally Equivalent to the EM Algorithm, *Journal of Royal Statistical Society* B, Vol. 57, 425-437.

[5] Little, R.J. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*, New York : Wiley.

[6] Tanner, M.A. (1991). Tools for Statistical Inference, *Lecture Notes in Statistics*, 67, Berger, S. *et al.*(Eds.), Springer-Verlag.