

4차 고다드 알고리즘의 몇 가지 수렴 성질: 2. 평균성능함수의 기하

正會員 최진호*, 배진수**, 송익호**, 박래홍***, 이민수**

Some Convergence Properties of Godard's Quartic Algorithm : 2. The Geometry of the Average Performance Function

Jinho Choi*, Jinsoo Bae**, Iickho Song**, Rae-Hong Park***,
Min Sou Lee** *Regular Members*

요 약

이 논문에서는 눈 먼 균등화에 쓰이는 4차 고다드 알고리즘의 수렴에 대하여 살펴보았다. 평균 성능 함수가 가장 작은 값을 갖는 점 부근에서 4차 고다드 알고리즘의 국소적 성질을 설명한 뒤, 평균 성능 함수의 기하에 대해 살펴보고 좋은 성질을 갖는 초기 매개변수 벡터는 평균 성능 함수의 기하를 써서 결정할 수 있음을 보였다.

ABSTRACT

Convergence analysis on Godard's quartic(GQ) algorithm used for blind equalization is accomplished in this paper. First, we explain the local behavior of the GQ algorithm around the global minimum point of the average performance function. Then we consider the geometry of the average performance function. The main result is that a good initial parameter vector of the GQ algorithm can be chosen based on the information of the geometry of the average performance function.

I. 머리말

학습 신호가 없을 때에도 쓸 수 있는 여러 가지 적응

여파 알고리즘들이 개발되어 왔는데 이들은 채널 균 등화나 신호 복원과 같은 곳에 쓸 수 있다. 눈 먼 균 등 화기도(blind equalization) 이런 적응 여파기의 한 갈 래이다. 적응 여파기의 출력을 $z(k)$ 라 하고 성능 함수

$$J(k) = \frac{1}{4} (z^2(k) - \gamma)^2 \quad (1)$$

*데이콤 종합연구소 PCS 연구팀
**한국과학기술원 전기 및 전자공학과
***서강대학교 전자공학과
論文番號:95265-0807
接受日字:1995年 8月 7日

을 생각해보자. 여기서, γ 는 입력 신호의 통계량으로 결정한다. 또한 기울기 경사법을 쓰는 여파기에서 매개변수는 다음 식을 써서 결정할 수 있다.

$$\theta(k+1) = \theta(k) - \mu \frac{\partial J(k)}{\partial \theta(k)} \quad (2)$$

여기서, μ 는 적응 이득이다. 이것이 눈 먼 균등화를 위한 4차 고다드 알고리즘이다. 이 알고리즘은 눈 먼 균등화에 많이 쓰인다.

이 논문에서는 1부에 [1] 이어 4차 고다드 알고리즘이 PAM신호의 눈 먼 균등화에 쓰일 때에 4차 고다드 알고리즘의 성질을, 곧, 4차 고다드 알고리즘의 평균 성능 함수의 기하를 살펴보았다. 이 기하를 쓰면 좋은 성질을 갖는 눈 먼 균등화기의 초기 매개변수 벡터를 얻을 수 있다.

II. 평균 성능 함수의 수렴속도와 안정성

먼저 1부에서 [1] 다룬 내용을 간단히 살펴보자. 잡음이 없을 때, 충격응답이 $\{h_i, i=0, 1, \dots\}$ 인 선형채널의 입력 $\{a(i), i=1, 2, \dots\}$ 이 독립이고 같은 확률을 가지는 PAM신호라면 그 출력은

$$y(k) = \sum_{i=0}^{\infty} h_i a(k-i) \quad (3)$$

이다. 균등화기의 매개변수 벡터가 $\theta(k) = [\theta_0(k), \theta_1(k), \dots]^T$ 라면 균등화기의 출력은

$$z(k) = \sum_{i=0}^{\infty} \theta_i(k) y(k-i) \quad (4)$$

이다. 적응 알고리즘 (2)를 쓸 때 γ 를

$$\gamma = \frac{E\{a^4(k)\}}{E\{a^2(k)\}} \quad (5)$$

로 잡고 매개변수 벡터 $\theta(k)$ 의 처음 값을 최적 매개변수 θ^* 과 충분히 가깝게 잡으면 부호 사이 간섭을 없애주는 벡터 θ^* 을 얻을 수 있다. 곧, $\theta(k) = \theta^*$ 이면 성능 함수 $J(k)$ 는 가장 작은 값을 갖게되고 시간 지연이 있을 수는 있지만 균등화기의 출력은 입력신호와 비슷하게 된다[2]. 채널에 잡음이 있다면 출력의 스펙트럼을 뭉툭하게 퍼져서 그 값의 정확도가 떨어지게 된다.

4차 고다드 알고리즘이 안정적이려면 신호 크기가 가질 수 있는 값들이 너무 퍼져 있으면 안된다. 보기를 들면, $a^2(k) \in \{1, 16, 256, 4096\}$ 이고 동일한 분포를 갖는다면 이 알고리즘은 불안정하게 된다. 그런데, 실제상황에서는 인접한 두 신호사이의 거리는 일반적으로 같으므로 대부분의 실제 신호에 대해 안정성을 보장할 수 있다.

적응 알고리즘에서는 적응 이득 μ 가 수렴속도와 안정도를 결정하는데, 성능 함수가 볼록 함수라 할지라도 μ 가 너무 크면 적응 알고리즘이 발산하므로 적응 이득 μ 의 상한값은 적응 알고리즘이 수렴하도록 정해져야 한다. 적응 알고리즘이 수렴하도록 하는 μ 의 범위는 다음과 같다.

$$0 < \mu < \frac{2}{\delta_3 \lambda_{\max}^G} \quad (6)$$

식 (6)에서 λ_{\max}^G 는 $\delta \equiv E^2[a^2(k)] - E[a^4(k)]$, $\tau = -3\delta h_0^2$, $\hat{\tau} = \tau/\delta_3$, $\delta_3 \equiv 3E^2[a^2(k)] - E[a^4(k)]$ 일 때 행렬 $R_h + \hat{\tau}I_0$ 의 가장 큰 고유값이며, 행렬 R_h 는 $[R_h]_{m+1, n+1} = \sum_{p=0}^{\infty} h_p h_{p+m-n}$ 으로 정의된다.

4차 고다드 알고리즘의 수렴속도를 결정 지향(decision directed) 적응 알고리즘과 견주어 보자[3, 4]. 결정지향 알고리즘의 성능 함수는

$$J_D(k) = \frac{1}{2} \{z(k) - dec(z(k))\}^2 \quad (7)$$

여기서, $dec(\cdot)$ 는 결정 장치이다. 이 알고리즘이 수렴하려면 다음 조건이 만족되어야 한다.

$$0 < \mu < \frac{2}{E\{a^2(k)\} \lambda_{\max}^D} \quad (8)$$

여기서, λ_{\max}^D 는 행렬 R_h 의 가장 큰 고유값이다.

평균성능함수 $\bar{J}_0 = E\{J(k)\}$ 의 헤시안 $M(\theta^*) = \frac{\partial^2}{\partial \theta^2} \bar{J}_0|_{\theta=\theta^*}$ 의 가장 큰 고유값 λ_{\max}^M 은 [5]

$$\lambda_{\max}^M = \delta_3 \lambda_{\max}^G \geq \delta_3 \lambda_{\max}^D \quad (9)$$

를 만족시킨다. 결정 지향 알고리즘과 4차 고다드 알

고리즘을 알맞게 정규화하면, 4차 고다드 알고리즘의 정규화된 적응 이득은 결정 지향 알고리즘의 정규화된 적응 이득보다 항상 작게 된다. 이의 특별한 경우가 [6]에서도 다루어졌다.

III. 4차 고다드 알고리즘의 평균 성능 함수의 기하

식 (2)의 평균을 얻으면

$$\theta(k+1) = \theta(k) - \mu \frac{\partial \bar{J}_0}{\partial \theta} \Big|_{\theta=\theta(k)} \quad (10)$$

임을 알 수 있다.

식 (10)의 매개변수 벡터 $\theta(k)$ 가 최적 매개변수 벡터 θ^* 와 비슷한 값을 갖는다면

$$\begin{aligned} \frac{\partial}{\partial \theta} \bar{J}_0 \Big|_{\theta=\theta(k)} &= \frac{\partial}{\partial \theta} \bar{J}_0 \Big|_{\theta=\theta^*} + M(\theta^*)(\theta(k) - \theta^*) \\ &\quad + O(\|\theta(k) - \theta^*\|^2). \end{aligned} \quad (11)$$

여기서, $\|\cdot\|$ 은 벡터의 크기를 나타낸다.

성능 함수가 가장 작은 값을 갖는 점 근처에 충분히 넓은 흡인 영역이 있다면 성능 함수가 가장 작은 값으로 수렴할 확률이 높을 것이다. 이 절에서는 4차 고다드 알고리즘의 평균 성능 함수의 기하와 흡인 영역을 살펴보았다. 식 (11)에서 최적 매개변수 벡터 θ^* 에 수렴하게 되는 매개변수 벡터 $\theta(k)$ 의 집합을 흡인 영역이라 한다.

어떤 함수가 볼록 집합 Ω 위에서 볼록함수이면 그 함수의 헤시안은 Ω 위에서 반양정치를 (positive semidefinite) 갖는다는 것과 그 역도 성립한다는 것이 알려져 있다 [7]. $\hat{\theta}$ 를 θ^* 와 충분히 가까운 매개변수 벡터라 하고 $\check{\theta} = \hat{\theta} - \theta^*$ 이라 하면

$$\begin{aligned} z^2(k) &= (\hat{\theta}^T Y(k))^2 \\ &= ((\theta^* + \check{\theta})^T Y(k))^2 \\ &\approx a^2(k) + 2a(k)(\check{\theta}^T Y(k)) + O(\|\check{\theta}\|^2) \end{aligned} \quad (12)$$

이고 헤시안은

$$\begin{aligned} M(\hat{\theta}) &= M(\theta^*) + 6E[a(k)Y(k)Y^T(k)(\check{\theta}^T Y(k))] \\ &\quad + O(\|\check{\theta}\|^2) \end{aligned} \quad (13)$$

이다. 여기서 $E[a(k)Y(k)Y^T(k)(\check{\theta}^T Y(k))]$ 를 $P(\check{\theta})$ 라 두자. 식 (13)의 $O(\|\check{\theta}\|^2)$ 을 무시하면, 매개변수 벡터 $\check{\theta}$ 가 흡인 영역 안에 있으려면

$$\lambda_{\min}(M(\check{\theta})) \geq 0, \quad (14)$$

곧,

$$\lambda_{\min}(M(\theta^*) + 6P(\check{\theta})) \geq 0. \quad (15)$$

여기서, $\lambda_{\min}(A)$ 는 행렬 A 의 가장 작은 고유값이다. 부등식 (15)에서 흡인 영역을 얻을 수 있다. 흡인영역의 정확한 범위를 알기는 어렵지만 흡인 영역 안에 있는 $\hat{\theta}$ 의 범위는 다음과 같이 얻을 수 있다.

$$\|\check{\theta}\| \leq \frac{\delta_3}{6} \cdot \frac{\lambda_{\min}^G}{\sqrt{E[\|a(k)Y(k)\|^2]E[\|Y(k)\|^4]}} \quad (16)$$

여기서, $O(\|\check{\theta}\|^2)$ 항은 생략했으며 λ_{\min}^G 는 행렬 $R_k + \tau I_0$ 의 가장 작은 고유값이다. 식 (16)으로 결정된 범위는 흡인 영역의 부분 집합이며 볼록 집합이다. 또 (16)을 근사하면 다음 식을 얻을 수 있다.

$$\|\check{\theta}\| \leq \frac{\delta_3}{6E[\|a(k)Y(k)\|]E[a^2(k)]} \cdot \frac{\lambda_{\min}^G}{\lambda_{\max}^D} \quad (17)$$

여기서, τ 가 작으면 $\lambda_{\min}^G/\lambda_{\max}^D \approx 1/S_D$ 이고 고유값 분포 S_D 는 흡인 영역의 반경에 영향을 미친다. 곧, S_D 가 작으면 흡인 영역의 반경은 커진다. 이것은 매우 중요한 결과이다. 왜냐하면 S_D 가 작으면 전력 스펙트럼은 평평한 모양을 갖는데, 일반적으로 평평한 모양의 전력 스펙트럼을 갖는 신호는 부호 사이 간섭이 심하지 않기 때문이다. 또, (16)과 (17)로부터 δ_3 가 커질수록 흡인 영역이 커져 더 많은 수의 초기 매개변수 벡터에 대해 수렴을 보장할 수 있게 된다. 그러나, δ_3 가 크려면 입력 신호가 간단해야 한다. 실제로, 1과 -1 두 값을 같은 확률로 갖는 2 PAM 신호의 δ_3 가 같은 전력을 갖는 N PAM 신호의 δ_3 가운데 가장 크다.

여기서, 볼록 영역과 오목 영역을 살펴 보자. 일반적으로 평균 성능 함수가 간단한 꼴이 아니기 때문에 이 두 영역을 정확히 얻을 수는 없으므로 이 두 영역의 부분 집합을 얻어보자. 먼저, 볼록 영역은 다음과 같이 정의된다.

$\Omega_{convex} \equiv \{M(\hat{\theta}) \text{이 반양정치인 } \hat{\theta}\}$

여기서,

$$\begin{aligned} M(\hat{\theta}) &= E[(3z^2(k) - \gamma)Y(k)Y^T(k)] \\ &= E[(3(\hat{\theta}^T Y(k))^2 - \gamma)Y(k)Y^T(k)]. \end{aligned}$$

이 볼록 영역은 볼록 집합은 아니다. 집합 Ω_{convex} 를 직접 얻을 수는 없지만, $\hat{\theta} \in \Omega_{convex}$ 일 충분조건은

$$\|\hat{\theta}\| \geq \sqrt{\frac{\gamma}{3} \cdot \frac{\lambda_{\max}^p}{(\lambda_{\min}^p)^2}} \quad (18)$$

이다. 식 (18)에서 정의된 Ω_{convex} 의 부분집합을 R_{convex} 라 하자. 곧,

$$R_{convex} = \{\hat{\theta} \mid \|\hat{\theta}\| \geq \sqrt{\frac{\gamma}{3} \cdot \frac{\lambda_{\max}^p}{(\lambda_{\min}^p)^2}}\}.$$

그러면, $R_{convex} \subset \Omega_{convex}$ 이므로 최적 매개변수 벡터 θ^* 이 언제나 R_{convex} 안에 있다고 말할 수는 없으나 최적 매개변수 벡터 θ^* 은 R_{convex} 경계 가까이에 있을 것이라 예상할 수 있다. 곧, 초기 매개변수 벡터는 R_{convex} 의 경계 위에서 고르면 된다.

한편, 오목 영역은 다음 식으로 정의된다.

$\Omega_{concav} \equiv \{M(\hat{\theta}) \text{이 반음정치인(negative semidefinite) } \hat{\theta}\}$.

볼록 영역에 대해서와 같은 방법으로 Ω_{concav} 의 부분 집합을 얻어보자.

$$\|\hat{\theta}\| \leq \sqrt{\frac{\gamma}{3} \cdot \frac{\lambda_{\min}^p}{E[\|Y(k)\|^4]}} \quad (19)$$

라면 $\hat{\theta} \in \Omega_{concav}$ 이 되며, Ω_{concav} 의 또 다른 부분 집합은 다음과 같이 얻을 수 있다.

$$\|\hat{\theta}\| \leq \sqrt{\frac{\gamma}{3E[\|Y(k)\|^2]}} \cdot \frac{\lambda_{\min}^p}{\lambda_{\max}^p}. \quad (20)$$

식 (20)에서 정의된 부분 집합은 원점을 포함하는 공이다. 이런 경우에는 매개변수 벡터 θ^* 은 늘 Ω_{concav} 밖에 있어 초기 매개변수 벡터를 (19)나 (20)에 정의된 영역 안에서 고를 필요는 없다.

평균 성능 함수의 헤시안은 볼록 영역에서는 음이

아니고 오목 영역에서는 양이 아니다. 따라서, 평균 성능 함수의 기하는 이들 영역에서는 매우 간단하다. 그러나, 이 두 영역 사이에서 평균 성능 함수의 기하는 그리 간단하지 않고 알아내기 매우 어렵다. 하지만, 고유값 분포 S_b 가 작은 값이면 이 영역도 작아짐을 (18)과 (20)에서 알 수 있다. 채널의 전력 스펙트럼이 평평한 꼴을 가져 고유값 분포 S_b 가 1과 가까운 값이면 평균 성능 함수의 기하는 간단해지고, 흡인 영역이 커지게 됨을 (17)에서 알 수 있다.

또 [8]에서처럼 채널-균등화가 조합 시스템을 이용해 평균 성능 함수의 기하를 얻을 수도 있으나, 균등화의 매개변수 벡터를 통해 얻은 평균 성능 함수의 기하에 대한 정보가 채널-균등화가 조합 시스템을 이용해 얻은 정보보다 초기 매개변수 벡터를 정하는데 더 쓸모있다. 왜냐하면, 채널-균등화가 조합 시스템은 채널을 알 때에만 쓸 수 있기 때문이다.

IV. 보 기

채널을 AR(1)이라고 하자. 곧,

$$g_0 y(k) + g_1 y(k-1) = a(k) \quad (21)$$

여기서, g_0 의 값은 모르고 입력 신호 $a(k)$ 는 독립이고 같은 분포를 갖고 있으며 $\{-3, -1, 1, 3\}$ 의 값을 갖는다고 하자. 적용 알고리즘

$$\begin{aligned} \theta(k+1) &= \theta(k) - \mu E\{(z^2(k) - \gamma)Y(k)Y^T(k)\} \theta(k) \\ &= \{I - \mu E[(z^2(k) - \gamma)Y(k)Y^T(k)]\} \theta(k) \end{aligned} \quad (22)$$

와 선형화된 적용 알고리즘

$$\tilde{\theta}(k+1) \simeq \tilde{\theta}(k) - \mu(3B_y - \gamma R_y) \tilde{\theta}(k) \quad (23)$$

의 매개변수 벡터의 궤적을 그림 1과 2에 보였다. 각 차분 방정식에 대한 적용 이득은 다음 식으로 결정하였다.

$$\mu = \frac{1}{2\text{tr} M(\theta^*)}.$$

이 그림들에서 원점은 최적 매개변수 벡터 $\theta^* = [g_0,$

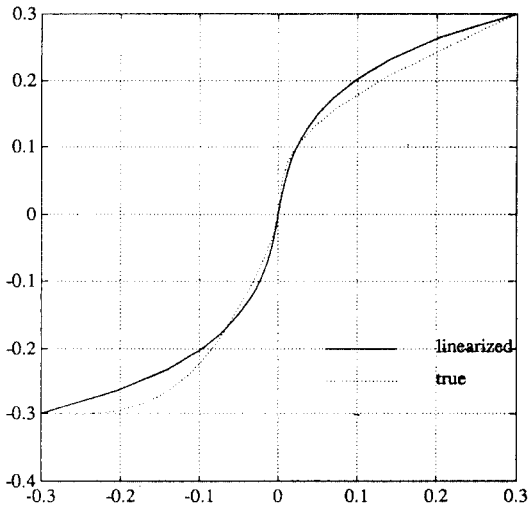


그림 1. 채널 (21)에서 알고리즘 (22)의 매개변수 벡터의 궤적과 (점선) 알고리즘 (23)의 매개변수 벡터의 궤적 (실선); $g_0 = 1, g_1 = 0.2$

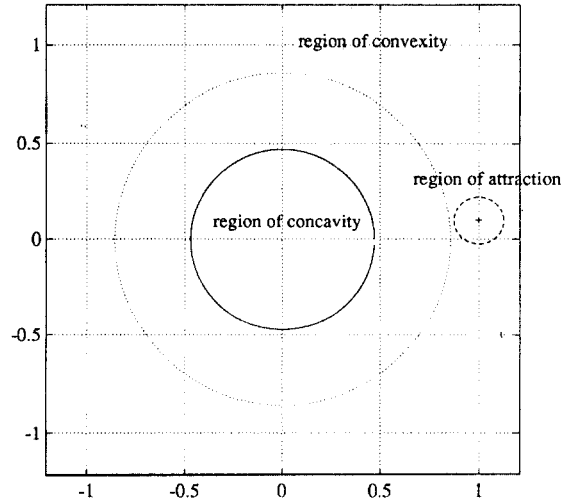


그림 3. 채널 (21)에서 평균 성능 함수의 기하; $g_0 = 1, g_1 = 0.1$

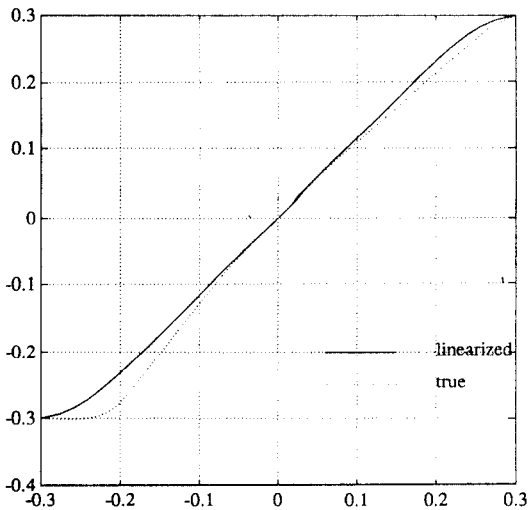


그림 2. 채널 (21)에서 알고리즘 (22)의 매개변수 벡터의 궤적과 (점선) 알고리즘 (23)의 매개변수 벡터의 궤적 (실선); $g_0 = 1, g_1 = 0.9$

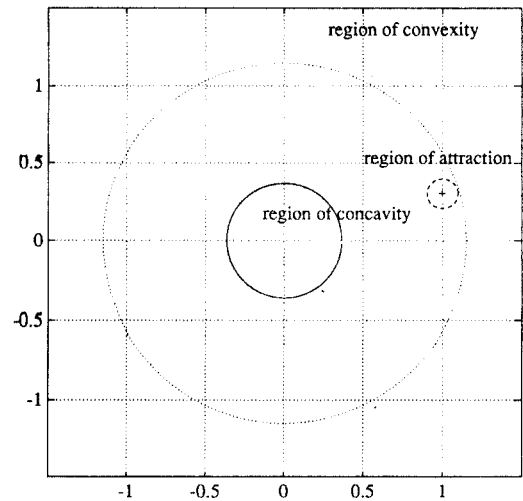


그림 4. 채널 (21)에서 평균 성능 함수의 기하; $g_0 = 1, g_1 = 0.3$

- 부 록 -

$g_1]^T$ 를 나타내며 초기 매개변수 벡터로는 $\check{\theta}(0) = \theta^* + [0.3, 0.3]^T$ 와 $\theta^* - [0.3, 0.3]^T$ 가 쓰였다. 그림에서 볼 수 있듯이 두 궤적은 그리 다르지 않다. 특히 원점 부근에서는 거의 같은 궤적을 나타내고 있다. 따라서, 최적 매개변수 벡터 θ^* 부근에서의 4차 고다드 알고리즘의 국소적 성질을 알려면 선형화된 균등화기를 쓸 수 있다. 그림 1과 2에 쓰인 채널에서 적응 이득과 흡인 영역 (16)의 반지름을 표 1에 보였다. 그림 1과 2의 두 초기 매개변수 벡터는 모두 (16)의 흡인 영역의 밖에 있지만 최적 매개변수 벡터로 수렴하므로 (16)의 흡인 영역은 실제 흡인 영역의 부분 집합임을 알 수 있다.

그림 3과 4에는 (17), (18), (20)의 부분집합들이 몇몇 채널에 대해 그려져 있다. 흡인 영역은 언제나 블록 영역 안에 있기는 하지만 블록 영역의 부분 집합 (18) 밖에 있을 수도 있다(그림 4). 균등화기 입력 $y(k)$ 의 공분산 행렬 고유값 분포가 작으면 흡인 영역은 커진다. 이 채널에서 $g_0=1$ 일 때 고유값 분포는 g_1 의 값을 따라 결정된다. 곧, 그림 3과 4에 나와 있듯이, $|g_1|$ 의 값이 커지면 고유값 분포도 커져서 흡인 영역은 작아지고 R_{convex} 의 반지름은 커짐을 알 수 있다.

표 1. 두 채널의 대한 적응 이득과 흡인 영역의 반지름

매개변수	적응이득 (μ)	흡인영역의 반지름
$g_0=1, g_1=0.2$	0.0042	0.1112
$g_0=1, g_1=0.9$	0.0012	0.0080

V. 맺음말

일반적으로 평균 성능 함수의 기하를 생각할 때 채널 균등화기 조합 시스템을 이용하는 것이 균등화기의 매개변수 벡터를 이용하는 것보다 간단하다. 그러나, 실제로는 매개변수 벡터를 이용해 얻은 성능 함수의 기하가 더 중요하다. 이 논문에서는 부분적이긴 하지만 성능 함수의 기하에 대한 쓸모있는 정보를 얻었다. 이는 때때로, 평균 성능 함수가 최소값이 아닌 극소값으로 수렴하는 것을 막아주곤 한다. 물론 이와 같은 부분적인 정보로는, 언제나 이런 극소값으로의 수렴을 늘 막을 수 있는 것은 아니다. 따라서 이런 극소값 수렴을 완벽하게 막을 수 있도록 많은 성능 함수의 기하에 대해 좀더 연구가 이루어져야 할 것이다.

먼저 [5]와 [9]에서 다음 부등식을 생각해 보자.

$$(F1) \lambda_{\min}(A+B) \geq \lambda_{\min}(A) + \lambda_{\min}(B), \quad \lambda_{\max}(A+B) \leq \lambda_{\max}(A) + \lambda_{\max}(B),$$

$$(F2) |x^T y| \leq \|x\| \|y\|,$$

$$(F3) E\{a|b|\} \leq E\{|a|\} E\{|b|\},$$

$$(F4) E\{|ab|\} \leq \sqrt{E\{|a|^2\}} \sqrt{E\{|b|^2\}},$$

여기서 $a, b \in \mathbb{R}, x, y \in \mathbb{R}^n, A, B \in \mathbb{R}^{n \times n}$ 이고 n 은 자연수이다.

(F1)을 쓰면, (15)에서 다음을 얻는다.

$$\lambda_{\min}(M(\theta^*) + 6P(\check{\theta})) \geq \lambda_{\min}(M(\theta^*)) - 6\lambda_{\max}(P(\check{\theta})). \quad (A.1)$$

그런데, $\|e\|=1$ 과 다음을 만족시키는 벡터 e 가 존재하므로,

$$\begin{aligned} \lambda_{\max}(E\{\check{\theta}^T a(k)Y(k)Y(k)Y^T(k)\}) \\ = E\{\check{\theta}^T a(k)Y(k)e^T Y(k)Y^T(k)e\}, \end{aligned} \quad (A.2)$$

(F2), (F3), (F4)를 쓰면 다음을 얻는다.

$$\begin{aligned} \lambda_{\max}(P(\check{\theta})) &= \lambda_{\max}(E\{\check{\theta}^T a(k)Y(k)Y(k)Y^T(k)\}) \\ &\leq E\{|\check{\theta}^T a(k)Y(k)| e^T Y(k)Y^T(k)e\} \\ &\leq \|\check{\theta}\| E\{\|a(k)Y(k)\| e^T Y(k)Y^T(k)e\} \\ &\leq \|\check{\theta}\| \sqrt{E\{\|a(k)Y(k)\|^2\}} \sqrt{E\{|e^T Y(k)Y^T(k)e|\}^4} \\ &\leq \|\check{\theta}\| \sqrt{E\{\|a(k)Y(k)\|^2\}} \sqrt{E\{\|Y(k)\|^4\}} \end{aligned} \quad (A.3)$$

끝으로 (A.1)과 (A.3)에서부터 (16)을 얻는다.

$$\|\check{\theta}\| \leq \frac{\lambda_{\min}(M(\theta^*))}{6\sqrt{E\{\|a(k)Y(k)\|^2\}} \sqrt{E\{\|Y(k)\|^4\}}} \quad (A.4)$$

마찬가지로, 근사식

$$\begin{aligned} &E\{\|a(k)Y(k)\| e^T Y(k)Y^T(k)e\} \\ &\approx E\{\|a(k)Y(k)\| E\{e^T Y(k)Y^T(k)e\}\} \\ &\leq E\{\|a(k)Y(k)\| \lambda_{\max}(E\{Y(k)Y^T(k)\})\} \end{aligned} \quad (A.5)$$

를 쓰면 (17)을 얻는다.

