

불일치된 최적 라플라스 양자기의 신호대잡음비 점근식의 유도

정회원 나 상 신*

Derivation of Asymptotic Formulas for the Signal-to-Noise Ratio of Mismatched Optimal Laplacian Quantizers

Sangsin Na* *Regular Member*

요 약

이 논문은 최소 평균제곱오차 라플라스 양자기가 평균이나 표준편차가 불일치된 신호에 적용될 때 야기되는 평균제곱오차 왜곡과 신호대 잡음비의 점근식을 유도한다. 이들 식은 양자점의 개수 N , 평균값의 변이량 μ , 양자기 설계 기준으로 사용된 표준편차에 대해 적용되는 신호의 표준편차 비율 ρ 로써 왜곡과 신호대잡음비의 직접적인 관계를 명확히 표시하고 있다. 수치 결과에 의하면, 논문의 주 근사식은, 효율 $R = \log_2 N$ 이 6 이상인 경우에, 상당히 넓은 μ 와 ρ 에 대해 신호대잡음비 참값의 1% 이내의 값을 예측하여 정확도가 아주 높은 것으로 판단된다. 이 논문을 통해 새로 발견된 점은 첫째 $\rho > 3/2$ 인 분산 강불일치의 경우에 신호대잡음비는 $9/\rho$ dB/bit 비율로 증가한다는 것과 둘째 최적 균일양자기는, 비록 최적으로 설계되었지만, 분산 임계불일치보다 조금 더 불일치된 것임을 밝힌 점이다. 또 μ 에 의한 신호대잡음비 손실은 비교적 크지 않은 것이 관찰되었다. 여기에 유도된 공식들은, 단구간 분산이 변하는 라플라스 분포로 잘 모형되는 음성이나 음악 신호를 하나의 양자기로 양자화하는 경우에 쓰임새가 있을 것으로 사료된다.

Key Words : Quantization, Mismatch, Laplace Density, Asymptotic Theory, Bennett's Integral

ABSTRACT

The paper derives asymptotic formulas for the MSE distortion and the signal-to-noise ratio of a mismatched fixed-rate minimum MSE Laplacian quantizer. These closed-form formulas are expressed in terms of the number N of quantization points, the mean displacement μ , and the ratio ρ of the standard deviation of the source to that for which the quantizer is optimally designed. Numerical results show that the principal formula is accurate in that, for rate $R = \log_2 N \geq 6$, it predicts signal-to-noise ratios within 1% of the true values for a wide range of μ and ρ . The new findings herein include the fact that, for heavy variance mismatch of $\rho > 3/2$, the signal-to-noise ratio increases at the rate of $9/\rho$ dB/bit, which is slower than the usual 6 dB/bit, and the fact that an optimal uniform quantizer, though optimally designed, is slightly more than critically mismatched to the source. It is also found that signal-to-noise ratio loss due to μ is moderate. The derived formulas can be useful in quantization of speech or music signals, which are modeled well as Laplacian sources and have changing short-term variances.

* 아주대학교 전자공학부 멀티미디어 신호처리 연구실 (sangna@ajou.ac.kr)

논문번호 : KICS2007-10-459, 접수일자 : 2007년 10월 8일, 최종논문접수일자 : 2008년 5월 2일

I. 서론

이 논문의 목적은, 라플라스 신호에 (최소 평균제곱오차의 의미에서) 최적으로 설계된 양자기가 평균 또는 분산이 다른 라플라스 신호에 적용될 때 발생하는, 평균제곱오차 왜곡과 신호대(양자화)잡음비의 공식을 유도하는 것이다. 고려되는 양자기는 고정요율을 사용하는 경우로 한정하였다. 이들 공식은, “불일치”된 양자기의 왜곡과 신호대잡음비를 양자점의 개수, 평균값의 변이량, 설계시에 사용된 신호원의 표준편차에 대해 적용되는 신호의 표준편차 비율의 직접적인 관계식으로서, 요율이 높아 갈수록 정확도가 증가하는 점근 공식들이다.

그림 1을 참조하여, 논문에 쓰일 용어와 양자기의 특성을 요약하면 다음과 같다. N 점 고정요율 홀양자기 Q_N 은, N 개의 실수값 양자점 $y_1 < y_2 < \dots < y_N$, ($x_1 = -\infty$ 과 $x_{N+1} = \infty$ 인) 경계값 $x_1 < x_2 < \dots < x_N < x_{N+1}$, 그리고 (중복하여 표기되는) 양자함수 Q_N 으로 완전히 표현되는데, 이 양자함수는 다대일 함수로서 $x \in (x_i, x_{i+1})$ 이면 $Q_N(x) = y_i$ 이다. 이 논문에서는 “고정요율”과 “홀양자기”만을 고려하므로, 편의상 이들 용어를 생략하고, 단순히 양자기라 부르기로 한다. 확률밀도 $q(x)$ 의 신호원에 최적인 N 점 양자기는 다음 평균제곱오차를 최소화하는 양자기를 말한다.

$$D(Q_N) = E_q \{ (X - Q_N(X))^2 \} \tag{1}$$

$$= \sum_{i=1}^N \int_{x_i}^{x_{i+1}} (x - y_i)^2 q(x) dx$$

평균 또는 분산 불일치된 양자화는 확률밀도 $q(x)$ 에 최적으로 설계된 양자기 Q_N 이 확률밀도 $p(x)$ 를 갖는 신호원에 사용되는 상황에서, $p(x)$ 와 $q(x)$ 는 예컨대 라플라스 밀도와 같은 동일한 종류의 확률밀도 함수에 속하지만, 평균값이나 분산이 다른 경우를 말한다. 이 논문에서는 $q(x)$ 와 $p(x)$ 가 라플라스 확률밀도 함수로서 다음 식으로 주어진 경우만을 대상으로 삼았다.

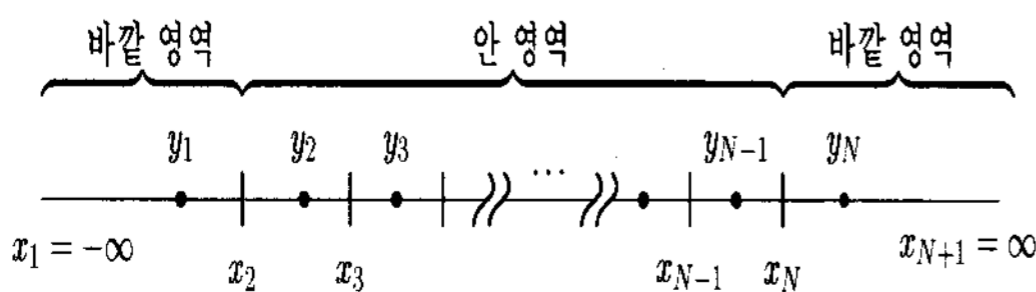


그림 1. 홀양자기의 표현

$$q(x) = \frac{1}{\sqrt{2}\sigma_q} \exp\left\{-\frac{\sqrt{2}|x|}{\sigma_q}\right\} \tag{2}$$

$$p(x) = \frac{1}{\sqrt{2}\sigma_p} \exp\left\{-\frac{\sqrt{2}|x-\mu|}{\sigma_p}\right\} \tag{3}$$

여기서 $\mu \geq 0$ 이라 가정하여도 일반성을 잃지 않으므로 이를 가정하였는데, 이 μ 는 평균의 불일치를 나타내는 지표로서 평균값의 변이량이다. 또 표준편차비 ρ 를 다음 식으로 정의하였는데

$$\rho = \frac{\sigma_p}{\sigma_q} \tag{4}$$

이 비율은 설계에 사용된 신호에 대하여 양자기에 적용되는 신호의 표준편차 비율로서, 궁극적으로 (표준편차의 제곱인) 분산의 불일치를 나타내는 지표로 쓰인다.

참고문헌 [1]에 따라, 그림 1에 보인 것처럼, 구간 $[x_2, x_N]$ 을 안 영역이라 부르며, 이 영역에 의한 왜곡을 안 왜곡이라 부르고 D_i 로 표기하기로 한다. 또 구간 (x_1, x_2) 와 (x_N, x_{N+1}) 의 합집합을 바깥 영역이라 부르며, 이에 의한 왜곡을 바깥 왜곡 D_o 로 표기한다. 최적 양자기 Q_N 의 경우에, 양자 경계값 x_N 또는 x_2 는 안 왜곡과 바깥 왜곡의 경계값이 되어 중요한 역할을 담당하므로, 중요 변수라 불린다.

양자기 불일치에 대한 최초의 연구는 안 왜곡 D_i 의 근사식을 제시한 Bennett [2]의 업적으로 보이는데, 이 식은 오늘날 베넷 적분식으로 불린다. 베넷 적분식은 예컨대 Bucklew [3]에 의해 엄정한 수렴식의 형태로 다차원으로 일반화되었고, Gray와 Linder [4]에 의해 엔트로피 제한된 가변요율 양자화로 확장되었다. 이들이 다루는 불일치 양자화는, 이 논문의 용어로 환언하면, $p(x)/q^{2/3}(x)$ 이 실수 집합 $(-\infty, \infty)$ 에 대해 적분가능하거나 $p(x)/q(x)$ 이 제한된 경우에만 의미를 갖는다. 이런 조건이 만족되는 경우에는, 바깥 왜곡이 안 왜곡에 비해 무시될 정도로 작기 때문에, 안 왜곡이 궁극적으로 전체 왜곡의 근사값으로 쓰일 수 있기 때문이다. 이런 이유로 말미암아, 이 조건은, 예컨대 [5,6]을 포함하여 양자화 불일치를 다루는 논문에서 사실상의 전제로 쓰이고 있다.

이와 달리, 이 논문에서는 이 조건을 전제하지 않는다. 이 논문의 진가는 오히려 $p(x)/q^{2/3}(x)$ 이 적분 불가능하거나 $p(x)/q(x)$ 이 무제한된 경우에 돋보인다고 할 수 있다. 이런 경우에는, 위 참고문헌들의

결과가 적용되지 않거나 무의미해지는데 반해서, 이 논문의 결과는 여전히 적용되기 때문이다. 이 논문에 의하여, 이 조건이 위배되는 때에 바깥 왜곡은 “보통의” $1/N^3$ 보다 서서히 감소하며, 불일치의 정도가 심한 경우에는 바깥 왜곡이 안 왜곡을 지배하게 된다는 것이 관찰되었다. 이런 경우 바깥 왜곡을 무시한다는 것은 큰 착오가 될 수 밖에 없을 것이다.

이 논문의 주요 결과는 왜곡과 신호대잡음비 공식이다. 이들 공식의 유도에 사용된 중요한 두 도구는 문헌 [1]의 중요 변수 x_N 의 점근 근사식과 [2]의 베넷 적분식이다. 유도된 공식은 양자점의 개수 N , 신호의 평균값 변이 μ , 그리고 표준편차비 ρ 를 사용하여 닫힌 형태로 표현되었다. 따라서 이들 공식은, 이들 요소가 왜곡에 영향을 끼치는 양뿐 아니라 그 구조까지를 명확히 밝힌다는 점에서, 그 가치가 크다고 할 것이다. (예컨대 실험수치 접근을 한 [7]의 내용을 설명하는 데 이 공식을 적용할 수 있다.)

유도된 공식의 정확도는 다음과 같다. 효율 $R = \log_2 N \geq 6$ 인 경우에, 이 논문의 주 공식 (17)은 광범위한 μ, ρ 값에 대해 신호대잡음비 참값의 1% 이내의 오차를 갖는다. 그리고 한번더 근사된 단항 근사공식도 충분한 정확도를 확보한 것으로 평가되는데, 구체적으로, 분산 약불일치 또는 강불일치 $\rho \neq 3/2$ 의 경우에, 효율 7 이상일 때, 참값과 오차 1% 이내의 신호대잡음비를 예측하며, 분산 임계불일치 $\rho = 3/2$ 의 경우에는 대략 5% 이내의 오차를 갖는 것이 관찰되었다. 물론 효율이 증가하면 근사식들의 정확도도 개선되는데, 이는 유도 과정에서 사용되는 기본식의 정확도 개선에 따른 당연한 귀결이다. 이들 근사식의 정확도를 받아들이는 경우에, 효율 한 비트당 증가하는 신호대잡음비는 ρ 값에 따라, 약불일치 $\rho < 3/2$ 경우에는 $6.02R$, 임계불일치 $\rho = 3/2$ 경우에는 $6.02R - 10 \log_{10} R$, 강불일치 $\rho > 3/2$ 경우에는 $9.03R/\rho$ dB임이 발견되었다. 이에 따라 예컨대, $\rho = 10$ 으로 심하게 표준편차가 불일치하는 경우에 $9.03/10 \approx 0.9$ dB/비트로 신호대잡음비가 증가하는데, 이는 통상의 6.02 dB/비트와 큰 차이가 난다. 또 임계불일치의 경우를 최적 균일양자화와 비교하면, 최적 균일양자기가 비록 신호원에 최적으로 설계되었지만, 최적 (비균일)양자기가 임계불일치된 경우보다 성능이 조금 더 떨어지는데, 이 사실이 근사 공식을 통해서 확인되었다. 이로써, 최적 균일양자기는 임계불일치된 양자기보다도 조금 더 분산 불일치된 것임을 밝혀냈다. 평균 불일치의 경우에, 평균값의 변이 μ 에 의한 성능저하는 상당히 완만함

이 관찰되었다.

이 논문의 체제는 다음과 같다. 제II절에서는 방법론을 설명·제시하고 왜곡을 위한 근사식을 유도한다. 제III절에서는 수치결과를 사용하여 주 공식과 이로부터 파생된 단항 근사식의 정확도를 평가한 뒤, 이들이 갖는 의미에 대해 논평한다. 끝으로 제IV절에서 요약 마무리한다.

II. 왜곡 근사식

라플라스 밀도함수 $q(x)$ 에 최적으로 설계된 N 점 양자기 Q_N 이 $p(x)$ 인 확률밀도함수를 갖는 신호에 적용되는 상황에서 양자기가 발생시키는 왜곡을 살펴보고자 한다. 양자기 Q_N 은 확률밀도 $q(x)$ 에 대한 최적 양자기이므로 다음 세 성질을 갖는다.

- (a) 경계값과 양자점의 대칭성: $i = 1, 2, \dots, N$,

$$x_i = -x_{N+2-i}, y_i = -y_{N+1-i} \quad (5)$$

- (b) 최외곽 양자점과 경계값의 상수 관계

$$x_2 - y_1 = y_N - x_N = \frac{\sigma_q}{\sqrt{2}} \quad (6)$$

- (c) 중요 변수 x_N 의 점근적 로그 증가 [1]

$$x_N \approx \frac{3\sigma_q}{\sqrt{2}} \ln \frac{N}{3} \quad (7)$$

이들은 논문의 전개 과정에서 사용된다.

방법론 왜곡 $D(Q_N)$ 의 근사식을 유도하는 데에 이 논문에서 사용한 방법은 다음과 같다. 첫째, 중요 변수 x_N 에는 식 (7)을 사용한다. 둘째, 안 왜곡 D_i 는 베넷 적분식으로 근사하고, 바깥 왜곡 D_o 는 이 논문에서 제안되는 근사식을 사용한다. 셋째, 위 두 근사식을 합하여 전체 왜곡을 근사한다.

평균제곱오차 왜곡 $D(Q_N)$ 은 안 왜곡 D_i 와 두 바깥 왜곡 $D_{o,l}, D_{o,r}$ 의 합 $D(Q_N) = D_i + D_{o,l} + D_{o,r}$ 으로 나타낼 수 있는데, 이들은 각각 다음과 같다.

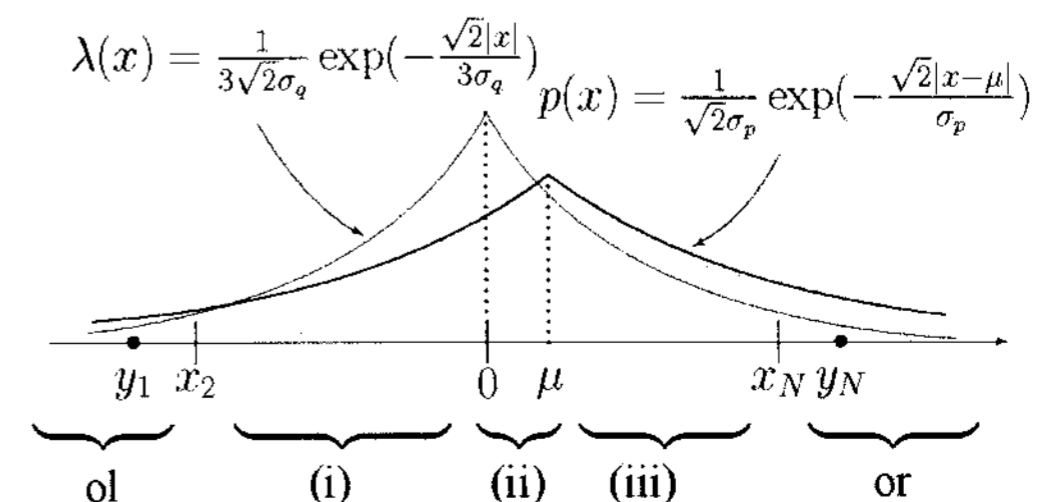


그림 2. 양자기 불일치를 위한 확률밀도 함수와 적분 영역

$$\begin{aligned}
 D_i &= \sum_{i=2}^{N-1} \int_{x_i}^{x_{i+1}} (x-y_i)^2 p(x) dx \\
 D_{o,l} &= \int_{x_1}^{x_2} (x-y_1)^2 p(x) dx \\
 D_{o,r} &= \int_{x_N}^{x_{N+1}} (x-y_N)^2 p(x) dx
 \end{aligned} \tag{8}$$

이제 이들 각각의 근사식을 유도한다.

2.1 안 왜곡의 근사식

안 왜곡 D_i 는 참고 문헌 [2]의 베넷 적분식 \tilde{D}_i 로 근사한다.

$$D_i \approx \tilde{D}_i = \frac{1}{12N^2} \int_{x_2}^{x_N} \frac{p(x)}{\lambda^2(x)} dx \tag{9}$$

여기서 $\lambda(x)$ 는 $q(x)$ 의 (점근적) 최적 양자점 밀도함수로서 $\lambda(x) = q^{1/3}(x) / \int_{-\infty}^{\infty} q^{1/3}(x) dx$ 이다. 이 식에 식 (2)의 $q(x)$ 를 대입하면 다음 식을 얻는다.

$$\lambda(x) = \frac{1}{3\sqrt{2}\sigma_q} \exp\left\{-\frac{\sqrt{2}|x|}{3\sigma_q}\right\}$$

함수 $p(x)$ 와 위 $\lambda(x)$ 를 대입하고 $\sigma_p = \rho\sigma_q$ 임을 사용하면

$$\tilde{D}_i = \frac{3\sqrt{2}\sigma_q}{4\rho N^2} \int_{x_2}^{x_N} \exp\left\{-\frac{\sqrt{2}(3|x-\mu|-2\rho|x|)}{3\rho\sigma_q}\right\} dx \tag{10}$$

이 적분은, 그림 2에 보인 대로, 적분 구간 $[x_2, x_N]$ 을 서로 소인 세 소구간 $[x_2, 0)$, $[0, \mu)$ 및 $[\mu, x_N]$ 으로 분리하여, 각 소구간에서 적분하여 더하면 되며, 구체적인 계산 결과는 다음과 같다.

1) $\rho \neq 3/2$ 인 경우 :

$$\begin{aligned}
 &\int_{x_2}^0 \exp\left\{-\frac{\sqrt{2}(3|x-\mu|-2\rho|x|)}{3\rho\sigma_q}\right\} dx \\
 &= \frac{3\rho\sigma_q}{\sqrt{2}(3-2\rho)} e^{-\sqrt{2}\mu/\rho\sigma_q} [1 - e^{(3-2\rho)\sqrt{2}x_2/3\rho\sigma_q}] \\
 &\int_0^\mu \exp\left\{-\frac{\sqrt{2}(3|x-\mu|-2\rho|x|)}{3\rho\sigma_q}\right\} dx \\
 &= \frac{3\rho\sigma_q}{\sqrt{2}(3+2\rho)} e^{-\sqrt{2}\mu/\rho\sigma_q} [e^{(3+2\rho)\sqrt{2}\mu/3\rho\sigma_q} - 1] \\
 &\int_\mu^{x_N} \exp\left\{-\frac{\sqrt{2}(3|x-\mu|-2\rho|x|)}{3\rho\sigma_q}\right\} dx \\
 &= \frac{3\rho\sigma_q}{\sqrt{2}(3-2\rho)} e^{\sqrt{2}\mu/\rho\sigma_q} \\
 &\quad \times [e^{-(3-2\rho)\sqrt{2}\mu/3\rho\sigma_q} - e^{-(3-2\rho)\sqrt{2}x_N/3\rho\sigma_q}]
 \end{aligned}$$

2) $\rho = 3/2$ 인 경우 :

$$\begin{aligned}
 &\int_{x_2}^0 \exp\left\{-\frac{\sqrt{2}(3|x-\mu|-2\rho|x|)}{3\rho\sigma_q}\right\} dx \\
 &= -e^{-2\sqrt{2}\mu/3\sigma_q} x_2 \\
 &\int_0^\mu \exp\left\{-\frac{\sqrt{2}(3|x-\mu|-2\rho|x|)}{3\rho\sigma_q}\right\} dx \\
 &= \frac{3\sigma_q}{4\sqrt{2}} e^{-2\sqrt{2}\mu/3\sigma_q} [e^{4\sqrt{2}\mu/3\sigma_q} - 1] \\
 &\int_\mu^{x_N} \exp\left\{-\frac{\sqrt{2}(3|x-\mu|-2\rho|x|)}{3\rho\sigma_q}\right\} dx \\
 &= e^{2\sqrt{2}\mu/3\sigma_q} (x_N - \mu)
 \end{aligned}$$

이 결과는 식 (10)으로부터 직접 계산하여 얻을 수도 있고, $\rho \neq 3/2$ 의 해당 식으로부터 $\rho \rightarrow 3/2$ 일 때의 극한값을 구하여 얻을 수도 있다.

이제 양자기의 경계값 대칭성 (5)와 x_N 의 점근식 (7)을 사용하면, D_i 의 근사식 \tilde{D}_i 를 얻는다. 곧

1) $\rho \neq 3/2$,

$$\tilde{D}_i \approx \frac{9}{2} \frac{\sigma_q^2}{N^2} \frac{e^{-\sqrt{2}\mu/\rho\sigma_q}}{3-2\rho} \left[\frac{2\rho + 3e^{(3+2\rho)\sqrt{2}\mu/3\rho\sigma_q}}{3+2\rho} - \frac{1 + e^{2\sqrt{2}\mu/\rho\sigma_q}}{2} \left(\frac{3}{N}\right)^{3/\rho-2} \right]$$

2) $\rho = 3/2$,

$$\tilde{D}_i \approx \frac{3\sigma_q^2}{N^2} \ln\left(\frac{N}{3}\right) \left[\cosh\left(\frac{2\sqrt{2}\mu}{3\sigma_q}\right) + \frac{1}{4\ln(N/3)} \sinh\left(\frac{2\sqrt{2}\mu}{3\sigma_q}\right) - \frac{\sqrt{2}}{6\ln(N/3)} \frac{\mu}{\sigma_q} e^{2\sqrt{2}\mu/3\sigma_q} \right] \tag{11}$$

위 식은, 정합 양자기의 경우, 곧 $\mu=0$ 이고 $\rho=1$ 인 경우에, [8]의 팬터 다윗 공식 $\tilde{D}_i \approx \frac{9}{2} \frac{\sigma_q^2}{N^2}$ 이 된다.

2.2 바깥 왜곡의 근사식

그림 2를 참조하여, 바깥 왜곡 D_o 는, 최외곽 구간 (x_1, x_2) 에 의한 왜곡 $D_{o,l}$ 과 최우측 구간 (x_N, x_{N+1}) 에 의한 왜곡 $D_{o,r}$ 의 합이다. 최외곽 양자점과 경계값의 관계식 (6)과 $p(x)$ 의 함수식을 사용하여 적분하면 다음을 어렵지 않게 구할 수 있다.

$$\begin{aligned}
 D_{o,l} &= \int_{-\infty}^{x_2} (x-y_1)^2 p(x) dx \\
 &= \frac{\sigma_q^2}{4} (2\rho^2 - 2\rho + 1) e^{-\sqrt{2}\mu/\rho\sigma_q} e^{\sqrt{2}x_2/\rho\sigma_q}
 \end{aligned} \tag{12}$$

$$D_{o,r} = \int_{x_N}^{\infty} (x - y_N)^2 p(x) dx$$

$$= \begin{cases} \frac{\sigma_q^2}{4} (2\rho^2 - 2\rho + 1) e^{\sqrt{2}\mu/\rho\sigma_q} e^{-\sqrt{2}x/\rho\sigma_q}, & 0 \leq \mu < x_N \\ -\frac{\sigma_q^2}{4} (2\rho^2 + 2\rho + 1) e^{-\sqrt{2}\mu/\rho\sigma_q} e^{\sqrt{2}x/\rho\sigma_q} \\ \quad + (\mu - y_N)^2 + \rho^2 \sigma_q^2, & \mu \geq x_N \end{cases} \quad (13)$$

$D_{o,r}$ 식은 평균값의 변이 μ 가 $0 \leq \mu < x_N$ 으로 안 영역에 있는지, 아니면 $\mu \geq x_N$ 으로 바깥 영역에 있는지에 따라서 적분결과가 달라져서, 위 (13)과 같이 표현된다. 이제 이를 합하여 총 바깥 왜곡 D_o 을 얻으면 다음과 같다.

$$D_o = \begin{cases} \frac{\sigma_q^2}{2} (2\rho^2 - 2\rho + 1) e^{-\sqrt{2}x/\rho\sigma_q} \cosh\left(\frac{\sqrt{2}\mu}{\rho\sigma_q}\right), & 0 \leq \mu < x_N \\ -\frac{\sigma_q^2}{2} e^{-\sqrt{2}\mu/\rho\sigma_q} \left[2\rho \cosh\left(\frac{\sqrt{2}x_N}{\rho\sigma_q}\right) \right. \\ \quad \left. + (2\rho^2 + 1) \sinh\left(\frac{\sqrt{2}x_N}{\rho\sigma_q}\right) \right] \\ \quad + (\mu - y_N)^2 + \rho^2 \sigma_q^2, & \mu \geq x_N \end{cases}$$

여기서, 지금까지의 바깥 왜곡 공식은 근사식이 아닌 정확한 식임을 덧붙인다. 이제 점근식 (7)을 x_N 에 대입하면, N 에 대한 바깥 왜곡 D_o 의 근사식 \tilde{D}_o 를 얻는다.

$$\tilde{D}_o = \begin{cases} \frac{\sigma_q^2}{2} (2\rho^2 - 2\rho + 1) \left(\frac{3}{N}\right)^{3/\rho} \cosh\left(\frac{\sqrt{2}\mu}{\rho\sigma_q}\right), & 0 \leq \mu < \frac{3\ln(N/3)}{\sqrt{2}}\sigma_q \\ -\frac{\sigma_q^2}{2} e^{-\sqrt{2}\mu/\rho\sigma_q} \left[2\rho \left(\frac{N}{3} + \frac{3}{N}\right) \right. \\ \quad \left. + (2\rho^2 + 1) \left(\frac{N}{3} - \frac{3}{N}\right) \right] \\ \quad + \sigma_q^2 \left[\left(\frac{\mu}{\sigma_q} - \frac{1}{\sqrt{2}} \ln(eN^3/27)\right)^2 + \rho^2 \right], & \mu \geq \frac{3\ln(N/3)}{\sqrt{2}}\sigma_q \end{cases} \quad (14)$$

다음 절에서 근사 공식의 정확도를 평가할 때는 식 (14)의 첫째 경우만을 대상으로 삼았는데, 큰 N 값의 경우에, 궁극적으로 $\mu < 3\sigma_q \ln(N/3)/\sqrt{2}$ 이 되는 이 경우가 논문의 관심사이기 때문이다.

2.3 총왜곡과 신호대잡음비

총왜곡 $D(Q_N)$ 의 점근 공식은 두 공식 (11)과 (14)의 합으로, \tilde{D} 라 표기하면, 다음 식을 얻는다.

$$\tilde{D} = \begin{cases} \frac{9}{2} \frac{\sigma_q^2}{N^2} \frac{e^{-\sqrt{2}\mu/\rho\sigma_q} (2\rho + 3e^{(3+2\rho)\sqrt{2}\mu/3\rho\sigma_q})}{3-2\rho} \\ \quad + \sigma_q^2 \left(\frac{3}{N}\right)^{3/\rho} \cosh\left(\frac{\sqrt{2}\mu}{\rho\sigma_q}\right) \frac{(\rho-1)^2(2\rho-1)}{2\rho-3}, & \rho \neq 3/2 \\ \sigma_q^2 \frac{3\ln(N/3)}{N^2} \cosh\left(\frac{2\sqrt{2}\mu}{3\sigma_q}\right) \\ \quad + \frac{3\sigma_q^2}{4N^2} \left[15 \cosh\left(\frac{2\sqrt{2}\mu}{3\sigma_q}\right) + \sinh\left(\frac{2\sqrt{2}\mu}{3\sigma_q}\right) \right] \\ \quad - \frac{2\sqrt{2}\mu}{3\sigma_q} e^{2\sqrt{2}\mu/3\sigma_q}, & \rho = 3/2 \end{cases} \quad (15)$$

식 (15)는 이 논문의 핵심으로서, 다음 절에서 이의 정확도를 평가하며 또 이로부터 더 간단한 부차적인 공식을 유도한다. 공식의 정확도는 실제로는 신호대잡음비 공식을 통해 평가하였는데, 이는 왜곡 값보다는 신호대잡음비가 양자기의 성능 지표로 더 흔히 사용되기 때문이다. 신호대잡음비는 다음과 같다.

$$\text{SNR} = 10 \log_{10} \frac{\sigma_p^2}{D(Q_N)} \quad [\text{dB}] \quad (16)$$

또 이것의 근사식 $\widetilde{\text{SNR}}$ 은 다음과 같이 계산하기로 한다.

$$\widetilde{\text{SNR}} = 10 \log_{10} \frac{\sigma_p^2}{\tilde{D}} \quad [\text{dB}] \quad (17)$$

III. 수치 결과

이 절에서는 최적 라플라스 양자기의 설계 과정을 설명하고, 앞에서 유도된 근사식의 정확도를 평가하여 공식의 타당성을 정립한 뒤, 이 공식들이 갖는 의미를 고찰한다.

3.1 최적 라플라스 양자기의 설계

평균 0, 표준편차 $\sigma_q = 1$ 인 라플라스 확률밀도 함수 $q(x)$ 에 최적의 양자기는 비반복 설계법에 의해 설계됐다. 이 설계법은 Nitadori가 [9]에서 최초로 제안했으며, 나중에는 [10,11] 등의 연구에서 사용됐다.

표 1. 분산 1인 라플라스 신호에 최적인 Q_N 의 특성

R	x_N	$D(Q_N)$	SNR
1	0.000000000000000	5.0000×10^{-1}	3.0103
2	1.12686252093771	1.7619×10^{-1}	7.5401
3	2.37958003705732	5.4476×10^{-2}	12.638
4	3.72396502364395	1.5373×10^{-2}	18.133
5	5.12587302714278	4.1022×10^{-3}	23.870
6	6.56044972900325	1.0610×10^{-3}	29.743
7	8.01250910028891	2.6987×10^{-4}	35.688
8	9.47362265931539	6.8062×10^{-5}	41.671
9	10.9393450773947	1.7090×10^{-5}	47.672
10	12.4073928332367	4.2821×10^{-6}	53.683
11	13.8766085449877	1.0717×10^{-6}	59.699
12	15.3464095586883	2.6807×10^{-7}	65.717
13	16.8165035426532	6.7037×10^{-8}	71.737
14	18.2867439007833	1.6761×10^{-8}	77.757
15	19.7570561962524	4.1907×10^{-9}	83.777
16	21.2273947808934	1.0477×10^{-9}	89.798

이 논문에서는 구체적으로 [12]에서 언급된 Lambert W 함수를 사용하여 수치의 정확도를 높였다.

표 1에 효율($R = \log_2 N$)의 범위 1~16에서의 최적 양자기 Q_N 의 중요 변수 x_N 과 신호대잡음비 참값 SNR이 수록됐다. 중요 변수는 예컨대 [1]에서 보고된 것과 잘 일치함이 관찰됐다. 여기에 수록되지 않았지만, $N^2 D(Q_N)$ 이 라플라스 신호원의 팬터

다잇 상수인 $9/2 = 4.5$ 로 수렴하는 추세도 잘 관찰되었다. 이들을 종합할 때, 양자기는 적절하게 설계되었으며, 불일치를 연구하는 데 적합한 것으로 평가된다.

3.2 주 근사식의 정확도 평가

설계된 양자기를 사용하여, 유도된 주 근사식 (17)의 정확도를 평가하였다. 이의 평가를 위해, 주 근사식의 신호대잡음비 \widehat{SNR} 을 참값 SNR과 비교하였다. 그림 3은 효율 $R=4 \sim 6$, 표준편차비 $\rho=0 \sim 3$, 평균값 변이 $\mu/\sigma_q = 0 \sim 1$ 의 곡면 그래프이다. 이 그림에서 \widehat{SNR} 의 곡면은, 참값 SNR보다 항상 작은 값이어서, 아래 쪽 곡면에 해당한다. 근사식의 정확도는 효율에 따라 증가하는데, 식 (11)과 (14)가 큰 N 에서 더 정확하게 성립하는 것에 따른 당연한 귀결이다. 구체적으로 $R=4$ 인 경우에 \widehat{SNR} 의 오차율은 5.2%, $R=5$ 의 경우에 1.8%, $R=6$ 의 경우에 0.92%, $R=7$ 의 경우에 0.6% 등을 보였다. 이 그림에서

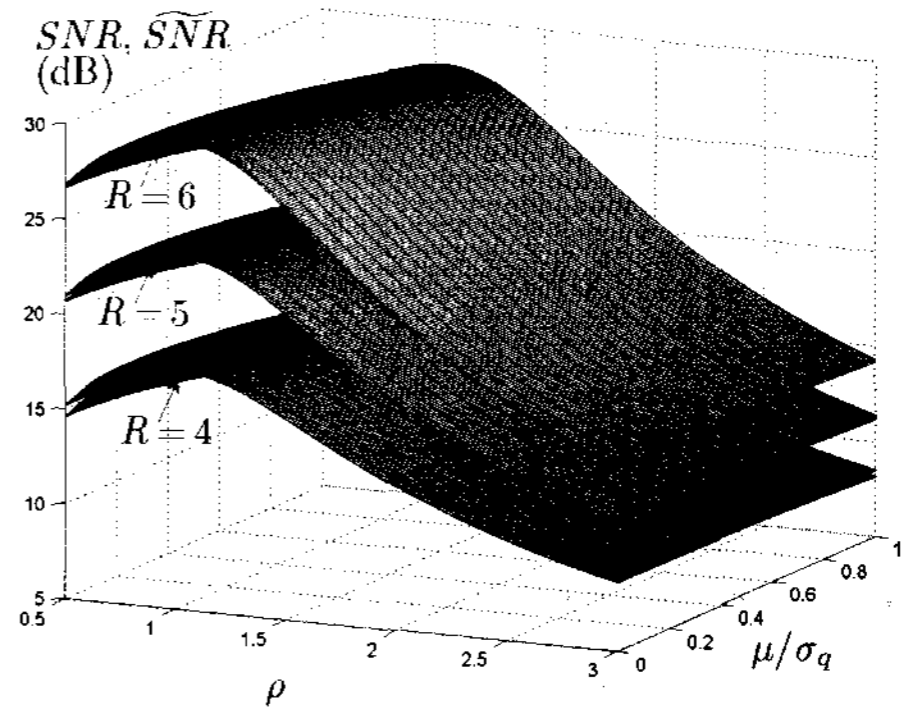


그림 3. SNR과 \widehat{SNR} 의 $\rho, \mu/\sigma_q$ 에 대한 곡면 그래프

$R=6$ 의 경우에 SNR과 \widehat{SNR} 의 값이 거의 같아 서로 겹쳐 보인다. 이 논문에는 실지 않았지만, 이런 추세는 $R \geq 7$ 경우에 더욱 강화됨이 관찰되었다.

참값 SNR과 근사값 \widehat{SNR} 의 대강은 언덕 모양을 하고 있다. N 과 μ 값이 고정된 경우에, 참값 SNR은 ρ 값이 1로 증가함에 따라 최대값으로 증가한 뒤, $\rho=1$ 이후에는 감소한다. 이 최대값은, $\rho=1$ 은 양자기와 신호원이 정합된 상태임을 감안할 때, 당연한 것이다. 한편 μ 값이 0으로부터 증가하면, SNR은 완만하게 단조감소한다. 이 언덕의 정점은 $\rho=1, \mu=0$ 에 위치하고 있으며, 바로 이 위치는 완전 정합된 (최적) 양자기의 성능을 얻는 점이다.

주 근사식 (15)에서 N 에 대해 가장 우세한 항 하나만으로 왜곡을 근사하는 식을 단항 근사식 D_d 라 하면, $\rho < 3/2$ 일 때는 식 (15)의 첫째식의 첫항을, $\rho = 3/2$ 일 때는 둘째식의 첫항을, 그리고 $\rho > 3/2$ 일 때는 첫째식의 둘째항을 뜻하게 된다. 이를 표 2의 단항 왜곡식으로 정리할 수 있다. 완전 정합 상태에서 단항식의 정확도를 평가한 결과(곧, 단항 근사식에 $\rho=1, \mu=0$ 을 대입하여 계산하면), 단항 근사식은 $R \geq 8$ 이면 참값과 오차율 1% 이내가 되어 높은 정확도를 갖는다.

표 2. 왜곡의 단항 근사식과 신호대잡음비 근사식

ρ	왜곡의 단항 근사식 D_d
$< 3/2$	$\frac{9 \sigma_q^2 e^{-\sqrt{2}\mu/\rho\sigma_q}}{2 N^2} \frac{2\rho+3e^{(3+2\rho)\sqrt{2}\mu/3\rho\sigma_q}}{3-2\rho} \frac{1}{3+2\rho}$
$3/2$	$\sigma_q^2 \frac{3 \ln(N/3)}{N^2} \cosh\left(\frac{2\sqrt{2}\mu}{3\sigma_q}\right)$
$> 3/2$	$\sigma_q^2 \left(\frac{3}{N}\right)^{3/\rho} \cosh\left(\frac{\sqrt{2}\mu}{\rho\sigma_q}\right) \frac{(\rho-1)^2(2\rho-1)}{2\rho-3}$

3.3 분산 불일치

분산 불일치는 평균은 일치하지만 표준편차는 달라 발생하는 불일치를 뜻한다. 이때는 $\mu=0$ 이어서, 근사식이 비교적 덜 복잡하다. 표 2의 단항 근사식 D_d 을 사용하여 $\mu=0$ 을 대입하면 얻는 신호대잡음비 근사식 \widehat{SNR}_1 을 사용하여 이의 정확도를 평가해 보기로 한다. 곧 $\widehat{SNR}_1 = 10 \log_{10} (\sigma_p^2 / D_d) |_{\mu=0}$ 이다. 구체적인 식은 다음과 같다.

$$\widehat{SNR}_1 = \begin{cases} 6.02R + 10 \log_{10} \frac{2\rho^2(3-2\rho)}{9}, & \rho < 3/2 \\ 6.02R - 10 \log_{10} R - 10 \log_{10} \frac{4 \ln 2}{3}, & \rho = 3/2 \\ \frac{9.03}{\rho} R + 10 \log_{10} \frac{\rho^2(2\rho-3)}{3^{3/\rho}(\rho-1)^2(2\rho-1)}, & \rho > 3/2 \end{cases} \quad (18)$$

이들 식은 각각 약불일치, 임계불일치, 그리고 강불일치에 해당한다. 그림 4에 ρ 에 대한 SNR 참값과 단항 근사식 \widehat{SNR}_1 의 값을 비교하였다.

3.3.1 약불일치

그림 4에 $\rho=1/2$ 인 경우를 약불일치의 대표적인 예로 보였다. 단항 근사식 \widehat{SNR}_1 은, $R \geq 6$ 이면, 1% 이내의 오차율을 갖는다. 이 경우는, $\rho=1$ 인 정합 여파기와 마찬가지로, 실질적으로 두 선이 포개져 한 선으로 보일 정도로 단항식의 정확도가 높다. 신호대잡음비가 6.02 dB/bit의 기울기로 증가한다는 사실이 그래프에서 잘 관찰된다.

“약불일치”라는 용어는, 불일치 손실이 적어 신호대잡음비가 정합 때와 증가율이 같다는 점에서 선택된 것이다. 그러나 염두에 두어야 할 사항이 있는데, 신호대잡음비가 ρ 에 따라 심하게 변할 수 있다는 사실이다. 단항 근사식 \widehat{SNR}_1 의 상수항 $10 \log_{10} (2\rho^2(3-2\rho)/9)$ 에 반영된 이 사실은 그림 5에 도시됐다. 예컨대 $\rho \approx 0$ 에서는 이 상수항으로 인한 신호대잡음비의 손실이 아주 크게 된다. 그것은 양자기의 설계에 사용된 표준편차에 비해 아주 작은 표준편차의 신호원이 양자화되는 경우에 실제로는 양자기의 중심 부분만 사용되고 나머지 부분은 “늘고” 있는—따라서 양자점의 개수에 비해 신호대잡음비는 낮은—상황이 전개되기 때문이다. (이와는 달리, 그림 5에서 $\rho \nearrow 3/2$ 인 경우는, 실제 신호대잡음비의 손실이 아니라, 임계불일치됨에 따라 야기되는 단항 근사식 \widehat{SNR}_1 의 부정확도에 의한 현상이다.) 그림 5에서 상수항 최대값은 $\rho=1$ 에서 발생하는데, 이것은 정합 상태에 해당한다.

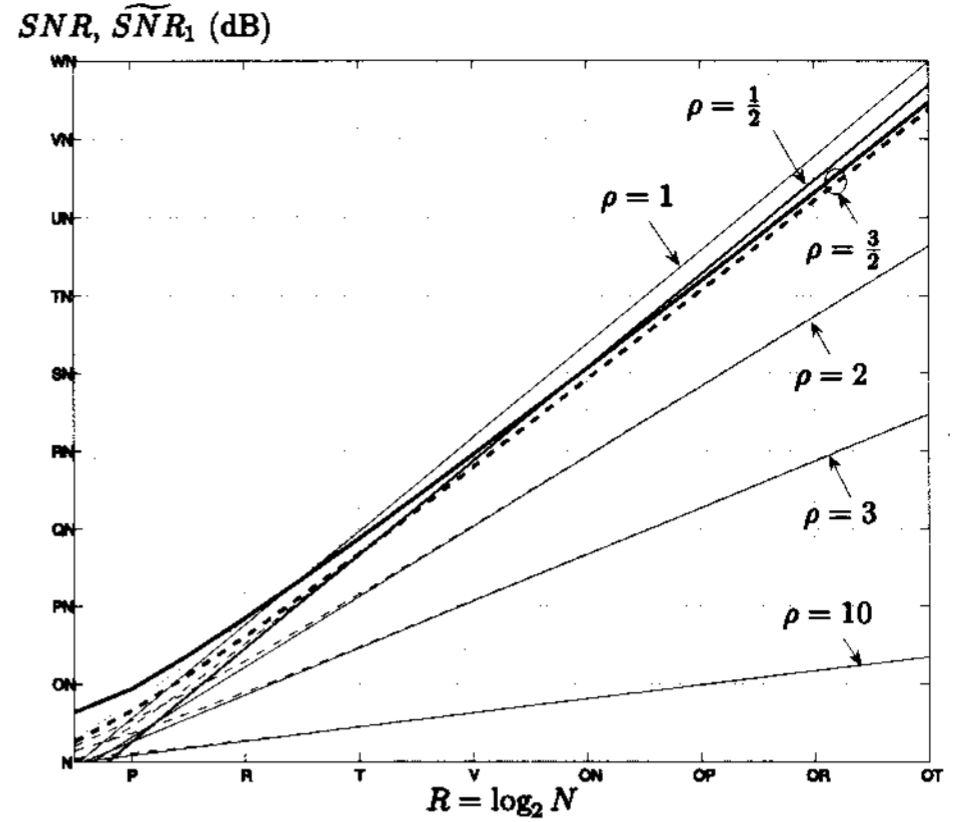


그림 4. 분산불일치된 양자기의 SNR(점선)과 단항 근사식 \widehat{SNR}_1 (실선)의 그래프

3.3.2 임계불일치

임계불일치는 $\rho=3/2$ 일 때의 불일치를 일컫는다. 임계불일치일 경우에는, $R \geq 8$ 일 때에, \widehat{SNR}_1 의 오차율은 5% 이내이다. 단항 근사식의 정확도가 약불일치나 (다음에 언급된) 강불일치일 경우보다 떨어진다는 사실은, 우세한 한 항으로는 왜곡을 잘 포착하지 못한다는 점을 반영한다. 그 이유는 식 (15)의 둘째항에 대해 우세한 첫항의 비율이 N 에 관한 로그 관계 $\ln(N/3)$ 을 갖는 때문일 것이다. 이것은 다른 경우의 지수 관계 $N^{2-3/\rho}$ 와 잘 대조된다. 물론 이 경우도 단항식의 정확도는 R 과 함께 좋아진다. (두 항을 사용하는 주 근사식 (17)의 \widehat{SNR} 의 경우에 오차율은 $R \geq 7$ 에서 0.4% 이내이다.)

신호대잡음비는 6.02 dB/bit의 증가율로 증가하는 것처럼 보인다. 그러나 단항 근사식 $\widehat{SNR}_1 = 6.02R - 10 \log_{10} R - 10 \log_{10} (4 \ln 2 / 3)$ 의 정확도를 인정하는 경우에, $-10 \log_{10} R$ 항으로 말미암아, 실제 증가율은 6.02 dB/bit보다 조금 작으며, 이는 그림 4에서도 관찰된다. 곧 $R=5$ 에서는 $\rho=1/2$ 경우보다 높은 곳에, $R \geq 12$ 일 때는 더 낮은 곳에 위치한다.

최적 (고정효율) 균일양자기의 신호대잡음비 점근식 $6.02R - 20 \log_{10} R + 10 \log_{10} (3/2 \ln^2 2)$ 이 참고문헌 [13]에 보고되어 있다. 이 식의 형태와 임계불일치 신호대잡음비 근사식 \widehat{SNR}_1 의 형태가 유사한데, 이들을 비교해 보기로 한다.

그림 6에 임계불일치된 최적 양자기와 최적 균일양자기의 신호대잡음비의 점근 근사식이, 참값 SNR과 주 근사식 \widehat{SNR} 와 함께, 도시됐다. 이를 비교하면, 신호대잡음비 측면에서, 최적 균일양자기는, 비록 신호원에 최적으로 설계되었지만 균일양자기라는

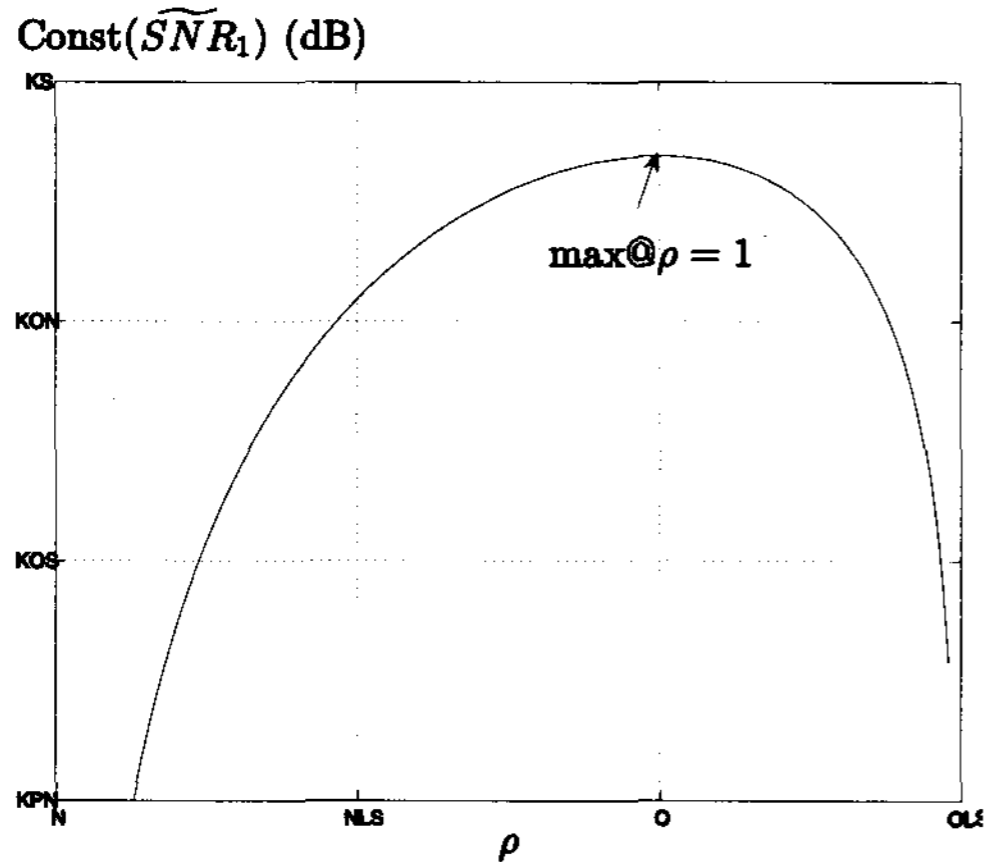


그림 5. ρ 에 대한 \widetilde{SNR}_1 의 상수항 변화

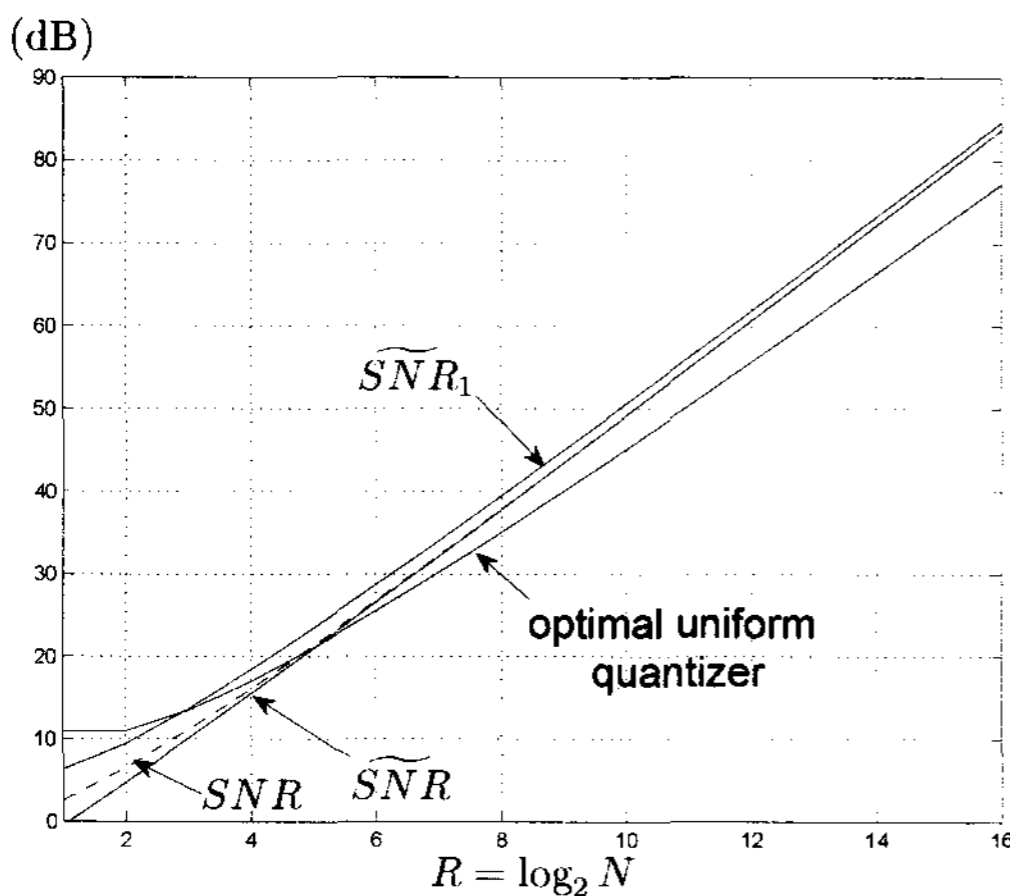


그림 6. 임계불일치된 최적 양자기와 최적 균일양자기의 성능 비교

한계 때문에, 최적 비균일 양자기가 임계불일치된 것보다 조금 더 불일치됨을 파악할 수 있다. 또 이 불일치에 의한 손실은, 임계불일치를 기준으로 할 때에, $10\log_{10}R - 10\log_{10}(2/\ln 2)$ 로서, 상당히 큰 값이 될 수 있다. 예컨대 $R=10$ 이면, 5.4 dB의 손실이 더 생긴다.

3.3.3 강불일치

강불일치란 $\rho > 3/2$ 일 때의 불일치를 일컫는다. 그림 4에서 $\rho=2, 3, 10$ 은 강불일치의 그래프이다. 효율 $R \geq 7$ 이면, \widetilde{SNR}_1 의 오차율은 1% 이내이다. 이 그래프는 기울기가 6.02 dB/bit보다 작으며, ρ 가 증가함에 따라 기울기는 더 작아진다. 근사식 \widetilde{SNR}_1 의 정확도를 인정하는 경우, 이 그래프들의 기울기는 $9.03/\rho$ dB/bit이다. 예컨대 $\rho=10$ 이면, $\widetilde{SNR}_1=0.903R-0.9993$ 을 얻

는데, 이 기울기 0.903은 정합 또는 약불일치 때의 6.02와 비교할 때, 아주 큰 차이가 난다.

3.3.4 신호대잡음비의 증가율

단항 근사식 \widetilde{SNR}_1 의 정확도가 확증된 것으로 가정하는 경우에, 표준편차비 ρ 에 의한 신호대잡음비 증가율의 영향을 고찰하기로 한다. 신호대잡음비 증가율은 식 (18)을 R 에 대해 미분하여 얻는데, ρ 가 0으로부터 3/2으로 증가하면, 6.02를 유지하다가 $\rho=3/2$ 에서 $6.02 - 20/(R\ln 10)$ 으로 갑자기 변화하는 불연속점을 갖는다. 또 $\rho > 3/2$ 에서는 $9.03/\rho$ 으로 ρ 에 반비례한다.

3.4 평균 불일치

평균 불일치란 $\rho=1$ 이고 $\mu \neq 0$ 인 경우에 발생하는 불일치를 말한다. 그림 7에 평균 불일치된 양자기의 신호대잡음비 SNR과 근사식 \widetilde{SNR}_2 를 도시하였다. 근사식 \widetilde{SNR}_2 는 식 (15)와 (17)로부터 $\rho=1$ 을 대입하여 얻으며, 구체적으로 다음과 같다.

$$\widetilde{SNR}_2 = 6.02R + 10\log_{10} \left[\frac{10}{9e^{-\sqrt{2}\mu/\sigma_q}(2 + 3e^{5\sqrt{2}\mu/3\sigma_q})} \right]$$

효율 $R \geq 6$ 이면, 두 그래프는 실질적으로 겹쳐 보이는데, 근사식이 1% 이내의 오차율을 갖기 때문이다. 평균 불일치에 의한 신호대잡음비 손실은 상대적으로 작은 것으로 판단된다. 예컨대 $R=8$ 일 때 손실은 $\mu/\sigma_q \leq 0.3$ 인 경우에 0.23 dB 이내이고, $\mu/\sigma_q=1$ 인 경우에 대략 2 dB 정도이다. 손실이 적은 이유는, 평균값의 변이가 크지 않은 경우에, 입력 신호 값 대부분이 양자기의 안 영역에서 처리되기 때문인 것으로 보인다.

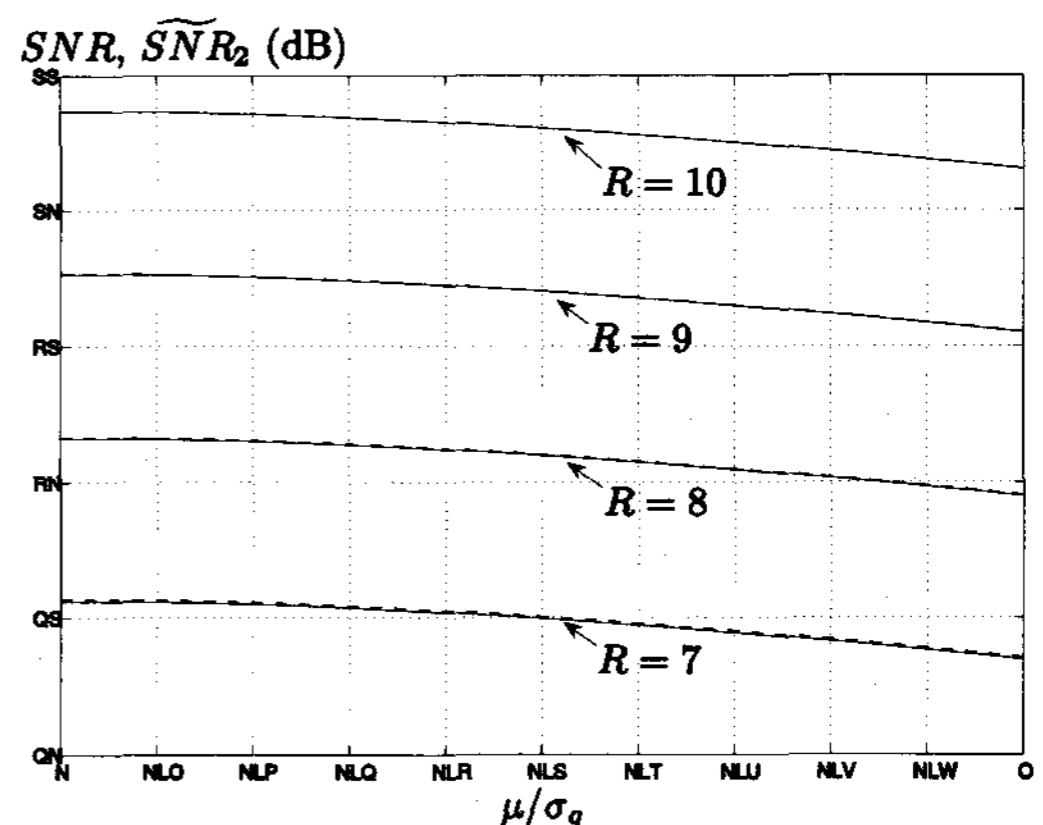


그림 7. 평균 불일치된 양자기의 신호대잡음비 SNR(점선)과 \widetilde{SNR}_2 (실선)

IV. 요약 및 결론

이 논문은 신호원에 불일치된 라플라스 양자기의 안 왜곡과 바깥 왜곡의 점근 근사식을 유도하여 이로부터 궁극적으로 신호대잡음비의 근사 공식을 유도하였다. 이들의 유도에는 양자기의 중요 변수의 점근식과 베넷 적분식이 가장 중요한 요소로 사용됐다. 유도된 주 공식은 효율이 7 이상인 경우에 충분한 정확도가 확보된 것으로 판정된다. 또 분산 불일치나 평균 불일치의 경우 모두, 단항 근사식이 유도됐는데, 이들도 높은 정확도를 갖는 것으로 관찰되었다. 이들 공식을 통해, 효율, 평균값의 변이, 분산 불일치의 정도가 신호대잡음비에 어떤 과정을 통해 그리고 얼마 정도의 영향을 끼치는지 명확히 예측할 수 있게 되었다. 이 논문에 의해 새롭게 발견된 사항은 첫째 $\rho > 3/2$ 인 분산 강불일치의 경우에 신호대잡음비는 비트당 $9.03/\rho$ dB 비율로 증가한다는 것과 둘째 최적 균일양자기는, 비록 최적으로 설계되긴 했지만, 분산 임계불일치된 최적 비균일 양자기의 성능보다 조금 더 떨어진다는 것이다. 또 평균의 변이 μ 에 의한 신호대잡음비 손실은 비교적 크지 않은 것이 관찰되었다. 여기에 유도된 공식들은, 음성이나 음악 신호를 하나의 양자기로 양자화하는 경우에 쓰임새가 있을 것으로 사료되는데, 이들 신호는 라플라스 분포로 잘 모형되며 또 단기간의 통계량이 심하게 변화하여 비적응 양자기의 경우 분산불일치를 야기할 수 있기 때문이다.

참 고 문 헌

[1] S. Na and D.L. Neuhoff, "On the support of MSE-optimal, fixed-rate, scalar quantizer," *IEEE Trans. Inform. Thy.*, Vol.IT-48, pp.239-247, Mar. 2002.

[2] W.R. Bennett, "Spectra of quantized signals," *Bell Syst. Tech. J.*, Vol.27, pp.446-472, July 1948.

[3] J.A. Bucklew, "Two results on the asymptotic performance of quantizers," *IEEE Trans. Inform. Thy.*, Vol.IT-30, pp.341-348, Mar. 1984.

[4] R.M. Gray and T. Linder, "Mismatch in high-rate entropy-constrained vector quantization," *IEEE Trans. Inform. Thy.*, Vol.IT-49, pp.1204-1217, May 2003.

[5] Y. Yamada, S. Tazaki, and R.M. Gray,

"Asymptotic performance of block quantizers with difference distortion measures," *IEEE Trans. Inform. Thy.*, Vol.IT-26, pp.6-14, Jan. 1980.

[6] Y. Yamada, S. Tazaki, M. Kasahara, T. Namekawa, "Varaince mismatch of vector quantizers," *IEEE Trans. Inform. Thy.*, Vol. IT-30, pp.104-107, Jan. 1984.

[7] W. Mauersberger, "Experimental results on the performance of mismatched quantizers," *IEEE Trans. Inform. Thy.*, Vol.IT-25, pp.381-386, Jul. 1979.

[8] P.F. Panter and W. Dite, "Quantization distortion in pulse count modulation with nonuniform spacing of levels," *Proc. IRE*, pp.44-48, Jan. 1951.

[9] K. Nitadori, "Statistical analysis of PCM," *Electron. Commun. in Japan*, Vol.-48, pp.17-26, Feb. 1965.

[10] H. Lanfer, "Maximum signal-to-noise-ratio quantization for Laplacian-distributed signals," *Information and System Theory in Digital Communications*, NTG-Report vol.-65, VDE-Verlag GmbH Berlin, Germany, p.52, 1978.

[11] P. Noll and Zelinski, "Comments on 'quantizing characteristics for signals having Laplacian amplitude probability density function'," *IEEE Trans. on Comm.*, Vol.COM-27, No.8, pp.1259-1260, Aug. 1979.

[12] G.J. Sullivan, "Efficient scalar quantization of exponential and Laplacian Randm Variables," *IEEE Trans. Inform. Thy.*, Vol.IT-42, pp.1365-1374, Sept. 1996.

[13] D. Hui and D.L. Neuhoff, "Asymptotic analysis of optimal fixed-rate uniform scalar quantization," *IEEE Trans. Inform. Thy.*, Vol.IT-47, pp.957-977, Mar. 2001.

나 상 신(Sangsin Na)

정회원

1991년 8월~현재 아주대학교 전
자공학부 재임
<관심분야> 자료압축, 정보이론
통신공학, 신호처리

