

두 가지 gradient 방법의 벡터 선형 예측기에 대한 적용 비교

신 광 균 양 승 인
 숭실대학교 전자 공학과

Comparison with two Gradient Methods through the application
 to the Vector Linear Predictor

Kwang Kyun Shin Seung In Yang
 Dept. of Electronic Eng., Soong Sil University

Abstract

Two gradient methods, steepest descent method and conjugate gradient descent method, are compared through application to vector linear predictors. It is found that the convergence rate of the conjugate gradient descent method is much faster than that of the steepest descent method.

1. 서론

일반적인 문제를 해결함에 있어 목적함수 $F(x)$ 를 정의하고 그 목적함수가 최소가 되도록 하는 x_0 를 구하는 경우가 많다. 이번 경우에 초기치를 가정한 후 반복적으로 원하는 해를 구하는 많은 하강 (Descent) 방법 중에서 간단하고 많이 알려진 steepest descent method (SDM)은 Gradient의 반대 방향으로 움직이는 것인데 수렴이 매우 느리게 되는 경우가 있다. 다른 방법으로는 Conjugate Gradient Descent Method (CGDM)가 있는데 다음과 같은 이유때문에 선택하였다 [1].

- 첫째로 이차수렴 성질을 갖는다.
 - n-차 Hilbert 공간에서의 이차 범 함수 (Quadratic Functional)에 대해서는 많아야 n번이면 수렴한다. 둘째, SDM에 비해 한 Step 당 계산 시간이나 기억용량이 조금 많을 뿐이다.
 - 셋째, 이론적으로 잘 규명되어 있다.
 - 넷째, 실제 사용해 보면 잘 동작한다.
- 앞에서 설명한 두가지 Gradient Method를 Vector Linear Predictor에 적용시켜 비교하고자 한다.

2. Conjugate Gradient Descent Method

여기서는 CGDM를 설명하고 SDM과 비교하고자 한다. CGDM에서는 x_i 을 초기치로, $P_i = -g_i$ 으로 놓은 후 x_i 로부터 다음해 x_{i+1} 는 다음과 같이 구해진다.

$$x_{i+1} = x_i + a_i \cdot P_i \quad \text{----- (1)}$$

여기서 그림 1 과 같이 P_i 는 이동방향이다. 또한 a_i 는 스칼라 상수이며 $F(x_i + a_i \cdot P_i)$ 을 최소화 시키도록 결정된다.

$$\begin{aligned} \text{즉 } \frac{d}{da} F(x_i + a \cdot P_i) \Big|_{a=a_i} &= 0 \\ &= F'(x_i + a_i \cdot P_i) \cdot P_i \\ &= \langle g(x_{i+1}), P_i \rangle \text{----- (2)} \\ &= 0 \end{aligned}$$

여기서 $\langle \quad \rangle$ 는 Scalar Product이며 $g(x_{i+1})$ 은 x_{i+1} 에서의 F의 Gradient이다.

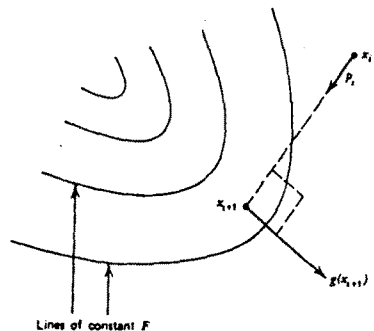


그림 1 CGDM에서의 x_i 와 x_{i+1} 의 관계
 fig. 1 The relationship of x_i and x_{i+1} in CGDM

식(2)가 의미하는 것은 X_{i+1} 에서의 Gradient 와 P_i 가 수직(Orthogonal) 이 되도록 X_{i+1} 을 결정 한다는 것이다 .

다음번 방향 P_{i+1} 은

$$P_{i+1} = -g_i + b_i \cdot P_i \quad \text{----- (3)}$$

$$b_i = \frac{\langle g_{i+1}, g_{i+1} \rangle}{\langle g_i, g_i \rangle} \quad \text{----- (4)}$$

여기서 g_i 는 X_i 에서의 F 의 Gradient 이다. SDM 는 식(3)에서 $b_i = 0$ 인 경우로서 방향이 항상 Gradient 의 반대 방향이 되는 경우이다.

3. Vector Linear Predictor 에 적용

Shannon 의 Rate - Distortion Theory 에 의하면 스키타 대신 벡터를 부호화 함이 더 좋은 성능을 주기 때문에 Vector Quantization 분야의 연구가 진행되고 있다 [2],[3],[4] .

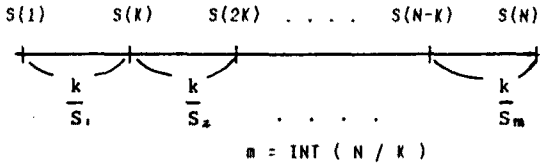


그림 2 벡터 \bar{S}_i 와 $S(i)$ 의 관계
fig. 2 The relationship of $S(i)$ and vector \bar{S}_i

그림 2와 같이 일단 신호 Sequence $S(i)$ 를 K개씩 Block 으로 쪼개서 각 BLOCK 을 Vector 라 생각하면 K - Dimensional Sequence \bar{S}_i 가 얻어진다.

과거의 P개 Vector $\bar{S}_{i-1}, \bar{S}_{i-2}, \dots, \bar{S}_{i-p}$ 보 부터 현재의 Vector 를 예측한 것을 \hat{S}_i 라 하면

$$\hat{S}_i = \sum_{j=1}^p a_j \cdot \bar{S}_{i-j} \quad \text{----- (5)}$$

여기서 a_j 는 앞으로 구하려는 스키타 상수로서 예측 계수라 한다. 최소화하려는 목적함수는 평균왜곡이며 다음과 같이 놓았다.

$$F(a) = \frac{1}{m-p-1} \cdot \sum_{i=p+1}^m \left\| \bar{S}_i - \hat{S}_i \right\|^2 \quad \text{----- (6)}$$

Simulation 에 사용된 Data Source 는 First - Order Gauss - Markov Source 로서 다음 차분방정식을 만족한다.

$$S_{n+1} = a \cdot S_n + W_n \quad \text{----- (7)}$$

여기서 a 는 Autoregression Constant 이며 $\{W_n\}$ 의 평균치는 0 , Variance 는 1 인 Gaussian Source 이다.

본 Simulation 에서는 a = 0.9 로서 매우 된 Correlated 경우만을 생각했다.

예측기 차수 P 는 6 으로 고정시켰으며, 예측계수의 초기치는 모두 0 으로 놓았다. Vector 의 Dimension k = 1 인 경우를 그림 3에, k = 3 인 경우를 그림 4 에 나타냈다 .

```

#####
k = 1
#####
a(1)  0.9407361895581271
a(2) -5.2691895344991323E-03
a(3)  2.5014043048545763E-03
a(4) -9.2427125267288332E-03
a(5)  5.6707580278290047E-03
a(6)  3.8907716410950336E-02
    
```

그림 3 - (a) 예측 계수 값 (k=1)
fig. 3 - (a) The value of Predictive Coefficient (k=1)

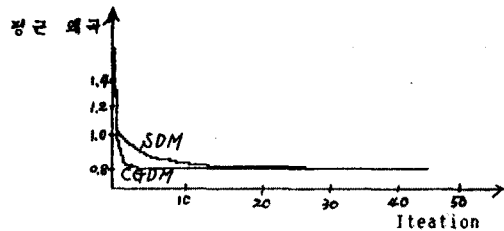


그림 3 - (b) CGDM 과 SDM 에 의한 평균 왜곡의 학습 곡선 (k=1)
fig. 3 - (b) Learning curves of the average distortion by CGDM and SDM (k=1)

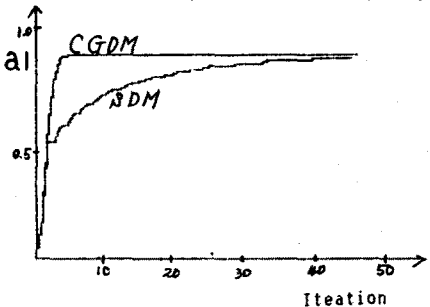


그림 3 - (c) CGDM 과 SDM 에 의한 예측 계수 a 의 비교 (k=1)
fig. 3 - (c) Learning curves of a by CGDM and SDM (k=1)

```

*****
K = 3
*****
a(1) 0.5638876941098054
a(2) 1.0973169818030808E-02
a(3) 2.7509858261774866E-02
a(4) 1.1949598427421981E-02
a(5) -4.5137598762684002E-02
a(6) 0.3900034425404388
    
```

그림 4 - (a) 예측 계수 값 (k=3)
fig. 4 - (a) The value of Predictive Coefficient (k=3)

두 경우 모두 그림 (a)에는 예측 계수들, 그림 (b)에는 반복 횟수에 대한 평균 왜곡의 변화를 그렸으며, 그림 (c)에는 첫번째 예측계수 a1의 각 Iteration에 의한 변화값을 그렸다.

앞의 3가지 그림을 비교하면 CGDM가 SDM보다 매우 빠르게 수렴하며, 많아야 예측기 차수 P 정도에서 수렴함을 알 수 있다.

평균 왜곡

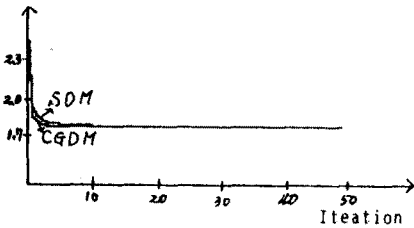


그림 4 - (b) CGDM과 SDM에 의한 평균 왜곡의 학습 곡선 (k=3)
fig. 4 - (b) Learning curves of the average distortion by CGDM and SDM (k=3)

4. 결론

본 논문에서는 여러가지 Gradient 방법 중에서 SDM과 CGDM을 선택하였으며, Micro-VAX SYSTEM을 사용하여 Vector Linear Predictor의 계수를 구하여 두 방법의 출력을 비교한 결과 CGDM가 SDM보다 수렴이 훨씬 빠르며 n-차원 Hilbert 공간에서는 최대한으로 n번만 반복하면 원하는 해를 구할 수 있다.

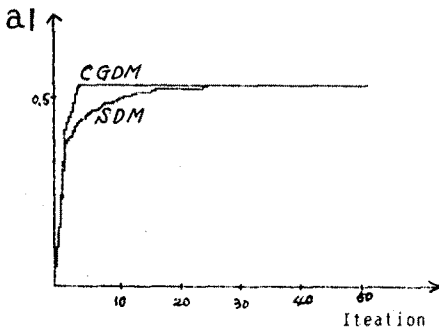


그림 4 - (c) CGDM과 SDM에 의한 예측 계수 a1의 비교 (k=3)
fig. 4 - (c) Learning curves of a1 by CGDM and SDM (k=3)

※ 참고 문헌 ※

[1] Lawrence Hasdorff, "Gradient Optimization and Nonlinear control", John Wiley & Sons, 1976.
[2] P.C.Chung & R.M.Geay, "Gradient Algorithms for Designing Predictive Vector Quantizers", IEEE Trans. ASSP - 34, no.4, pp 679 - 690, August 1986.
[3] Huseyin Abut, R.M.Gray & Guillermo Rebolledo, "Vector Quantization of Speech and Speech-Like Wave Forms", IEEE Trans. ASSP-30, no.3, pp423 - 435, June 1982.
[4] Vladimir Cuperman & Allen Gersho, "Vector Predictive Coding of speech at 16 kbit/s", IEE E Trans.COM - 33 NO.7 pp 685 - 696, July 1985.