

## 문제 4: Matrix Decomposition 기반 Multivariate Random Number 생성 병렬화 (대학원: 30점)

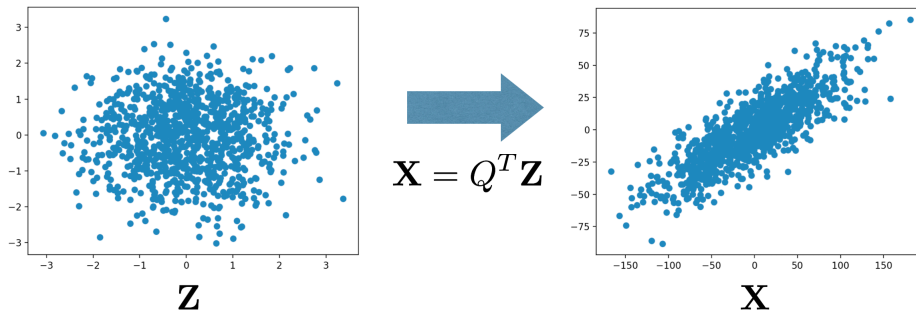
### 문제 배경

Random Number 생성은 몬테카를로 기법 등 확률론적 시뮬레이션(Stochastic Simulation)에서 필수적인 프로세스 중 하나로, 주어진 확률분포  $p(\mathbf{x})$  를 따르는 확률벡터열(Random Vector Sequence)  $\{\mathbf{X}_i\}_{i=1}^N$  을 생성하는 것이 목적이다.

$$\mathbf{X}_1, \dots, \mathbf{X}_N \sim p(\mathbf{x})$$

확률분포  $p(\mathbf{x})$  를  $d$  -차원 다변량정규분포(Multivariate Normal Distribution)  $\mathcal{N}(\mu, \Sigma)$  로 가정하는 경우,  $d$  -차원 표준정규분포  $\mathcal{N}(0, I_d)$  로부터 추출된 확률벡터열  $\{\mathbf{Z}_i\}_{i=1}^N$  로부터  $\{\mathbf{X}_i\}_{i=1}^N$  을 생성할 수 있다.  $(d \times d)$  -Matrix  $\Sigma$  를 Cholesky Decomposition  $\Sigma = QQ^T$  으로 분해했을 때,  $\{\mathbf{X}_i\}_{i=1}^N$  의 생성은 수식 (1)을 통한 변환 문제가 된다.

$$\mathbf{X}_i = \mu + Q^T \mathbf{Z}_i, \quad \forall i = 1, \dots, N \quad (1)$$



### 문제 설명

확률분포  $p(\mathbf{x})$  가  $\mu = 0$  인  $d$  -차원 다변량정규분포  $p(\mathbf{x}) = \mathcal{N}(0, \Sigma)$  를 따른다. 주어진 Cholesky Decomposition 알고리즘을 이용하여 Covariance Matrix  $\Sigma$  를  $QQ^T$  로 분해한 후 주어진 행렬  $\mathbf{Z} = [\mathbf{Z}_1, \dots, \mathbf{Z}_N]$  로부터 수식 (1)을 통해 행렬  $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_N]$  을 생성하는 순차코드가 C 와 Fortran 으로 주어져 있다.

주어진 순차코드를 병렬화하되, 다음에 제시된 병렬화 과정에서의 유의사항을 위반하는 경우 제출한 코드의 성능과 상관없이 0점으로 평가됨에 유의한다.

## Cholesky Decomposition Algorithm

- Input  $\Sigma$  :  $(d \times d)$  - Matrix
- Output  $Q$  :  $(d \times d)$  - Matrix

We want to derive the matrix  $Q$  where  $\Sigma = QQ^T$ . First set  $Q = 0$ .

```
loop for  $i = 1, \dots, d$ 
  loop for  $j = 1, \dots, i$ 
    set  $s = \sum_{k=1}^{j-1} q_{ik}q_{jk}$ 
    if  $i = j$ :
       $q_{ij} = \sqrt{\Sigma_{ii} - s}$ 
    else:
       $q_{ij} = \frac{\Sigma_{ij} - s}{q_{jj}}$ 
  end loop
end loop
```

## 참고 및 유의 사항

1. 본 문제는 확률/통계학 관련 사전지식이 없어도 풀 수 있음을 참고한다. Cholesky decomposition 및 Matrix multiplication 문제이며 행렬  $\mathbf{Z} = [\mathbf{Z}_1, \dots, \mathbf{Z}_N]$  는 확률변수가 아닌 고정된 숫자로 제시되었고 주어진 순차코드에 모든 정보가 주어져 있다. 제시된 Cholesky decomposition 알고리즘은 기본적인 선형대수 지식만 알면 이해할 수 있다.
2. 제시된 Cholesky decomposition 알고리즘의 Flow 변경은 가능하나 Random number 생성 알고리즘이나 또는 공개코드를 사용은 금지한다.
3. 소스 행렬의 분할이 없는 병렬화는 인정하지 않으며, 순차코드의 실행시간보다 3배 이상 시간이 더 걸리는 답안도 인정하지 않는다. 기타 이슈가 발생할 경우에는 출제자의 의도를 따른다.

## 평가 방법

1. 병렬코드가 Tachyon-II의 4노드(32 core) 에서 수행되는 시간을 time command로 측정한다. 순차코드와 함께 주어진 serial.sh 파일을 참고한다.
2. 순차코드 실행이 끝나면 result 폴더 밑에 result.txt 파일이 생성된다. 병렬 코드는 순차코드의 해와 오차범위내 ( $1e-6$ ) 에서 같은 해를 생성해야 하며, 본 조건이 만족되지 않는 경우 코드의 정확성 미달로 결격 처리한다.