Spectral State Compression - Algorithm 1

Annie Ye, yuhanye

Key Points:

- **Rank Reduction:** This method focuses on rank reduction by using the largest singular values, allowing the transition matrix to be low-rank, even if the empirical matrix isn't.
- **Data Compression:** The method still performs data compression, which is particularly important when dealing with high-dimensional transition matrices.
- **Projection to the Simplex:** The method projects the frequency estimates onto the probability simplex, ensuring that the rows of $\hat{P}$ remain valid probability distributions.

1. **Input:** A trajectory of observed states $\{X_1, \ldots, X_n\}$ and the known rank r, which is the target rank for the low-rank transition matrix.

2. **Construct Empirical Matrices:** From the observed trajectory, the empirical frequency matrix $\hat{F}$ is constructed. This matrix encodes the frequency with which transitions between different states occur, based on the sample path. $\hat{P}$ is the normalized version of this matrix, representing transition probabilities between states.

3. **Singular Value Decomposition (SVD):** Apply SVD to $\hat{F}$, decomposing it into three matrices:

$$\hat{F} = U_\Sigma \Sigma_F V_F^\top$$

   where U_Σ and V_F are orthogonal matrices and Σ_F is a diagonal matrix of singular values.

4. **Frequency Estimation:** The next step is to compute $\hat{F}_0$, which is the top-r rank approximation of $\hat{F}$. This is done by selecting the largest r singular values and the corresponding singular vectors. This helps in reducing the dimensionality while preserving the most significant transitions.

5. **Estimating the Transition Matrix:** Using the frequency matrix $\hat{F}_0$, the estimated transition matrix $\hat{P}$ is derived. Specifically, the entries of $\hat{P}$ are calculated as:

$$\hat{P}_{i,\cdot} = \frac{\hat{F}_{i,\cdot}}{\sum_{j=1}^{p} \hat{F}_{i,j}} \quad \text{if} \sum_{j=1}^{p} \hat{F}_{i,j} > 0$$

   and

$$\hat{P}_{i,\cdot} = \frac{1}{p} 1_p^\top \quad \text{if} \sum_{j=1}^{p} \hat{F}_{i,j} = 0$$

   This normalization ensures that each row of $\hat{P}$ sums to 1, preserving the properties of a Markov transition matrix.

The output of the algorithm consists of $\hat{P}$, the estimated low-rank transition matrix, and $\hat{F}$, the empirical frequency matrix.