Preconditioning for Stochastic Automata Networks

Amy Langville

Preconditioning 2001 Conference
Lake Tahoe, CA
April 29, 2001

Outline

- SANs
- Literature Review
- NKP Preconditioner: applied to general matrix, multilinear algebra
- NKP Preconditioner for SAN's Q matrix in descriptor form
- Implementation issues
- Suggestions

Markov chains

- a stochastic process which follows the Markov property. The state of the system at time t only depends on the most recent past.

- DTMC

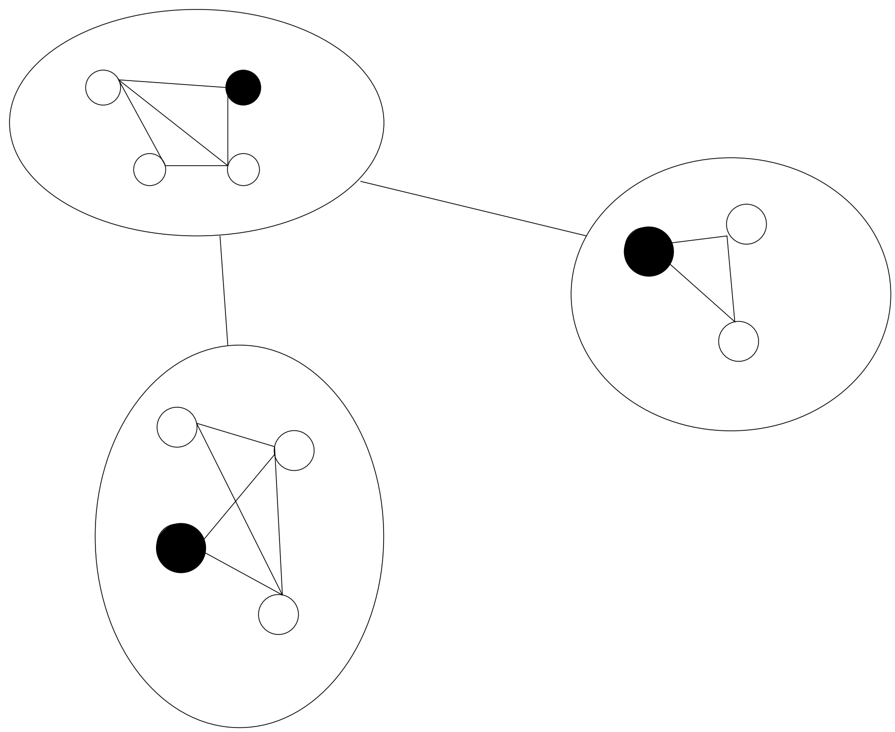
$$P = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n,1} & p_{n,2} & \cdots & p_{n,n} \end{pmatrix},$$

where $p_{i,j}$ holds the conditional probability of moving from state i to state j in one time step.

- CTMC

$$Q = \begin{pmatrix} q_{1,1} & q_{1,2} & \cdots & q_{1,n} \\ q_{2,1} & q_{2,2} & \cdots & q_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{n,1} & q_{n,2} & \cdots & q_{n,n} \end{pmatrix},$$

where $q_{i,j}$ holds the transition rate of moving from state i to state j and $q_{i,i} = -\Sigma_{j \neq i} q_{i,j}$.



Stochastic Automata Networks (SANs)

- Problem with MC analysis = size of P or Q.
- One solution = SANs (represent Q in compact form)
- A SAN is a collection of stochastic automata which act more or less independently requiring only infrequent interaction.

Tensor Algebra: $\otimes, \oplus$ are key operations in SAN formalism.

- $A \in \Re^{m_1 \times n_1}$, $B \in \Re^{m_2 \times n_2}$, then $A \otimes B \in \Re^{m_1 m_2 \times n_1 n_2}$ and is defined by

$$A \otimes B = \begin{pmatrix} a_{1,1}B & \cdots & a_{1,n_1}B \\ \vdots & \ddots & \vdots \\ a_{m_1,1}B & \cdots & a_{m_1,n_1}B \end{pmatrix}.$$

- $A \oplus B = A \otimes I_{m_2} + I_{m_1} \otimes B$. ($\oplus$ is defined in terms of $\otimes$.)

  (defined only for square matrices)

- Some Tensor Properties:
  - $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.
  - $\oplus_{i=1}^{N} A^{(i)} = \Sigma_{i=1}^{N} I_{n_1} \otimes \cdots \otimes I_{n_{i-1}} \otimes A^{(i)} \otimes I_{n_{i+1}} \otimes \cdots \otimes I_{n_N}$.

SAN descriptor

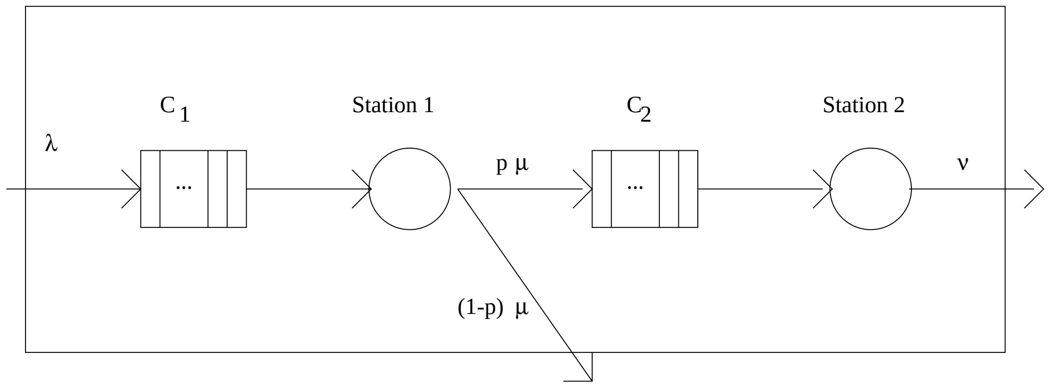
- MC has global Q which can be stored in much more compact form as the ordinary sum of Kronecker products of much smaller matrices.

$$Q = \sum_{j=1}^{2E+N} \otimes_{i=1}^{N} Q_j^{(i)},$$

where

N = number of automata,

E = number of synchronizing events.



SAN Example: A Simple Queueing Network

Consider a queueing network with two exponential, finite-capacity, single-server stations in tandem. When a station is full, customers are lost. For example, for $C_1 = 1$ and $C_2 = 2$, the transition rate matrix is given by

$$Q = \begin{pmatrix} -\lambda & 0 & 0 & \lambda & 0 & 0 \\ \nu & -(\nu+\lambda) & 0 & 0 & \lambda & 0 \\ 0 & \nu & -(\nu+\lambda) & 0 & 0 & \lambda \\ \mu(1-p) & \mu p & 0 & -\mu & 0 & 0 \\ 0 & \mu(1-p) & \mu p & \nu & -(\mu+\nu) & 0 \\ 0 & 0 & \mu(1-p) & 0 & \nu & -(\nu+\mu(1-p)) \end{pmatrix}.$$

Writing Q as a SAN

- In general, a SAN represents the global generator matrix as

$$Q = \oplus_{i=1}^{N} Q_l^{(i)} + \sum_{j=1}^{E} Q_{e_j},$$

where

N = number of automata,

E = number of synchronizing events,

$Q_l^{(i)}$ = automata i's local transition matrix,

Q_{e_j} = synchronizing event j's contribution to the global Q.

SAN Example: Writing Q as a SAN

- Our example has only 1 synchronizing event; a customer leaves station 1 and enters station 2. $A^{(1)}$ = station 1's automata and $A^{(2)}$ = station 2's automata. Thus,

$$Q = Q_l^{(1)} \oplus Q_l^{(2)} + Q_{e_1}$$

$$= \begin{pmatrix} \lambda & \lambda \\ \mu(1-p) & -\mu(1-p) \end{pmatrix} \oplus \begin{pmatrix} 0 & 0 & 0 \\ \nu & -\nu & 0 \\ 0 & \nu & -\nu \end{pmatrix} +$$

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & \mu p & 0 \\ 0 & 0 & \mu p \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} -\mu p & 0 & 0 \\ 0 & -\mu p & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

- We only need to store $Q_l^{(1)}$, $Q_l^{(2)}$, and the 4 matrices making up Q_{e_1}. For very large MCs, this is a huge savings in storage. The Global Q never needs to be formed or stored.

Stationary analysis of a SAN

GOAL: Solve the following system for π

$$\pi \left(\sum_{j=1}^{2E+N} \otimes_{i=1}^{N} Q_j^{(i)} \right) = 0,$$

$$\pi e = 1.$$

- Review of methods employed
  - Direct: L, U factors hard to get
  - Iterative: Power–Stewart. Splittings–Dayar.
  - Projection: Arnoldi, GMRES–Stewart. BiCGSTAB, QMR–Buchholz.
  - Decompositional: hierarchical SANs–Buchholz, NCD SANs–Dayar.
- Problem: SANs $\downarrow$ storage, $\uparrow$ complexity $\Rightarrow$ time til convergence=high.
- Solutions:
  - RSS - Ciardo, Miner, Kemper
  - Equivalent SANs - Buchholz
  - Preconditioners - Stewart, Buchholz, Chan

Existing SAN Preconditioners

- ILU Preconditioners are hard to employ.
- Stewart's Neumann series inverse: M^{-1} approximates $Q^{\#}$.

$$M^{-1} = \Sigma_{m=0}^{M} P^m.$$

- Buchholz's preconditioner, similar to Stewart's.
- Additive/Multiplicative Schwartz
- Diagonal: $M^{-1} = D^{-1}$, where D is the diagonal of Q.

Potential SAN Preconditioner: NKP ???

NKP Preconditioner for general matrix R

- Find A, B so that $R \approx A \otimes B$.

  Precondition with $M^{-1} = A^{-1} \otimes B^{-1}$.

- Pitsianis and Van Loan find A, B so that $\|R - A \otimes B\|_F^2$ is minimized by transforming this problem to an equivalent, easier problem.

- They show $\|R - A \otimes B\|_F^2 = \|\tilde{R} - ab^T\|_F^2$, where $\tilde{R}$ is a special rearrangement of the matrix R, $a = vec(A)$ and $b = vec(B)$.

Solving NKP Problem: $\min_{a,b} \|\tilde{R} - ab^T\|_F^2$.

- Rank-1 approximation problem has SVD solution.

$$a^* = \sigma_1 u_1, b^* = v_1.$$

- Reversing the vectorizing operation gives the optimal A, B matrices.

Advantages of NKP Preconditioner:

- To find A, B, only need $\tilde{R}$-vector multiplication algorithm.

  $\tilde{R}$ does not need to be formed or stored.

- Only little matrices A^{-1}, B^{-1} need to be formed and stored.

  Efficient vector-Kronecker product multiplication algorithm can be used so that M^{-1} is not explicitly needed.

The NKP, $A \otimes B$, when $R = \Sigma_{i=1}^{p} G_i \otimes F_i$

- $\tilde{R} = \Sigma_{i=1}^{p} g_i f_i^T$ where $g_i = vec(G_i)$ and $f_i = vec(F_i)$.

- It can be proven that when R has this special structure, the optimal A, B matrices are linear combinations of the matrices making up R.

$$
\begin{aligned}
A &= \alpha_1 G_1 + \alpha_2 G_2 + \cdots + \alpha_p G_p, \\
B &= \beta_1 F_1 + \beta_2 F_2 + \cdots + \beta_p F_p.
\end{aligned}
$$

- Original NKP problem can be written as a nonlinear optimization problem. Choose $\alpha_1, \alpha_2, \ldots, \alpha_p, \beta_1, \beta_2, \ldots, \beta_p$ so that this nonlinear convex function is minimized

$$
\|R - A \otimes B\|_F^2 = \| \sum_{i=1}^{p} G_i \otimes F_i - (\sum_{i=1}^{p} \alpha_i G_i) \otimes (\sum_{i=1}^{p} \beta_i F_i)\|_F^2
$$

$$
= \sum_{i=1}^{n_G} \sum_{j=1}^{n_F} \left[\left(\sum_{k=1}^{p} \sigma_{G_k}(i) \cdot \sigma_{F_k}(j) \right) - \left(\sum_{k=1}^{p} \alpha_k \sigma_{G_k}(i) \right) \left(\sum_{k=1}^{p} \beta_k \sigma_{F_k}(i) \right) \right]^2 .
$$

Back to SANs

- $Q = \Sigma_{j=1}^{2E+N} \otimes_{i=1}^{N} Q_j^{(i)}$ has special structure similar to $R = \Sigma_{i=1}^{p} G_i \otimes F_i$.

  But Q is Kronecker product of N terms, not just 2 terms.

- SAN Problem: find $A, B, \ldots, N$ to min $\|Q - A \otimes B \otimes \cdots \otimes N\|_F^2$.

  Precondition with $M^{-1} = A^{-1} \otimes B^{-1} \otimes \cdots \otimes N^{-1}$.

- Questions:

  – Finding A, B was based on the matrix SVD.

    Extension for finding $A, B, \ldots, N$?

  – When $R = \Sigma_{i=1}^{p} G_i \otimes F_i$, then $A = \Sigma_{i=1}^{p} \alpha_i G_i$, $B = \Sigma_{i=1}^{p} \beta_i F_i$.

    Extension to case when $R = Q = \Sigma_{j=1}^{2E+N} \otimes_{i=1}^{N} Q_j^{(i)}$?

  – To find A, B, we rearranged R to form $\tilde{R}$. For $A, B, \ldots, N$, what is the proper rearrangement of Q?

Multilinear Algebra

- The tensor is fundamental object of multilinear algebra.
  - 1^{st}-order tensor is a vector.
  - 2^{nd}-order tensor is a matrix.
  - 3^{rd}-order tensor is a 3D box.
- Operations on Tensors:
  - inner product
  - outer product
  - scalar product
  - tensor-matrix multiplication
- Outer product of N vectors $(a \circ b \circ \cdots \circ n)$ results in an N^{th}-order tensor.

The HOSVD

- Tensors have rank as well as a SVD, called the HOSVD.

- A rank-1 N^{th}-order tensor is written as the outer product of N vectors.

- HOSVD of N^{th}-order tensor R shows that R can be written as the sum of rank-1 tensors.

$$R = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} s_{i_1,i_2,\cdots,i_N} u_{i_1}^{(1)} \circ u_{i_2}^{(2)} \circ \cdots \circ u_{i_N}^{(N)}.$$

Multilinear Algebra and the NKP

- We want to min $\|R - A \otimes B \otimes \cdots \otimes N\|_F^2$ for a general matrix R.
- Following the 2D case, we need to show $\exists \tilde{R} \ni$

$$\|R - A \otimes B \otimes \cdots \otimes N\|_F^2 = \|\tilde{R} - a \circ b \circ \cdots \circ n\|_F^2.$$

- $a \circ b \circ \cdots \circ n$ is a rank-1 N^{th}-order tensor

  $\Rightarrow$ rearrangement of R ($\tilde{R}$) must be N^{th}-order tensor.

- Goal: approximate N^{th}-order tensor with rank-1 N^{th}-order tensor

  $\Rightarrow$ use truncated HOSVD.

Question: How do we define this N^{th}-order tensor $\tilde{R}$?

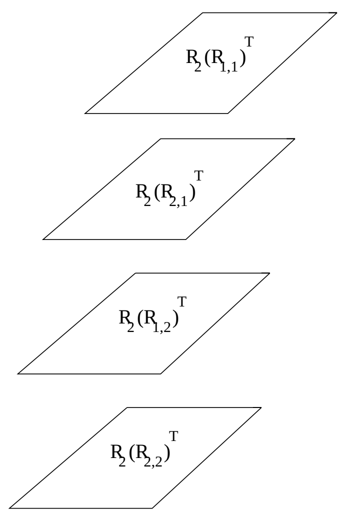
Defining Rearrangement Operator for Higher Orders

- For 2^{nd}-order tensor R, Pitsianis and Van Loan have already defined $\tilde{R} = R_2(R)$.
- The 3^{rd}-order rearrangement operator, $R_3(R)$, uses the 2^{nd}-order rearrangement operator.
- In general, we define the rearrangement operator recursively.

  We can then prove that with this recursive definition, the sums of squares match.

Example: Approximate $R_{12\times 12}$ by $A_{2\times 2}, B_{2\times 2}, C_{3\times 3}$.

$$
\begin{aligned}
R_{12\times 12} &= \begin{pmatrix} R_{1,1} & R_{1,2} \\ & \\ R_{2,1} & R_{2,2} \end{pmatrix}. \\
\tilde{R}_{4\times 4\times 9} &= R_3(R) =
\end{aligned}
$$



and then

$$\|R - A \otimes B \otimes C\|_F^2 = \|R_3(R) - a \circ b \circ c\|_F^2.$$

Approximating N^{th}-order tensor by a rank-1 N^{th}-order tensor

We want to min $\|\tilde{R} - a \circ b \circ \cdots \circ n\|_F^2$. where $\tilde{R}$ is an N^{th}-order tensor and $a \circ b \circ \cdots \circ n$ is a rank-1 N^{th}-order tensor.

- Unfortunately, truncating the HOSVD does not give the optimal rank-1 approximation to the N^{th}-order tensor $\tilde{R}$.
- de Lathauwer:
  - truncated HOSVD is *good* approximation.
  - HO power algorithm to find *optimal* rank-1 approximation.
- We are only looking for an approximate inverse preconditioner $\Rightarrow$ try truncating the HOSVD. Let

$$a = s_{1,1,\ldots,1} u_1^{(1)}, b = u_1^{(2)}, \ldots, n = u_1^{(N)}.$$

We hope $A \otimes B \otimes \cdots \otimes N$ is a good approximation to R.

Practical Issues

- We'd rather not form the N^{th}-order tensor $\tilde{R}$ or find its HOSVD.

  That'd be counter to the purpose of using a SAN.

- $\Rightarrow$ try to extend Pitsianis and Van Loan's special structure work.

  – Recall: $R = \Sigma_{i=1}^{p} G_i \otimes F_i,$

  $$\Rightarrow \text{optimal } A = \Sigma_{i=1}^{p} \alpha_i G_i, \text{ optimal } B = \Sigma_{i=1}^{p} \beta_i F_i.$$

  – NKP Problem transformed into a nonlinear convex optimization problem, where $\tilde{R}$ and the HOSVD of $\tilde{R}$ were never needed.

Question: Does this extend to Q's special structure?

Solving $\| \Sigma_{j=1}^{T} \otimes_{i=1}^{N} Q_j^{(i)} - A \otimes B \otimes \cdots \otimes N \|_F^2$, where $T = 2E + N$.

- We can show

$$
\begin{aligned}
A &\approx \alpha_1 Q_1^{(1)} + \alpha_2 Q_2^{(1)} + \cdots + \alpha_T Q_T^{(1)} \\
B &\approx \beta_1 Q_1^{(2)} + \beta_2 Q_2^{(2)} + \cdots + \beta_T Q_T^{(2)} \\
&\vdots \\
N &\approx \eta_1 Q_1^{(N)} + \eta_2 Q_2^{(N)} + \cdots + \eta_T Q_T^{(N)}.
\end{aligned}
$$

- Thus, just as in the optimal 2-matrix case, the original problem

$$\min \|Q - A \otimes B \otimes \cdots \otimes N\|_F^2$$

can be transformed into a nonlinear convex optimization problem.

$$\|Q - A \otimes B \otimes \cdots \otimes N\|_F^2 \approx \| \sum_{j=1}^{T} \otimes_{i=1}^{N} Q_j^{(i)} - (\sum_{j=1}^{T} \alpha_j Q_j^{(1)}) \otimes (\sum_{j=1}^{T} \beta_j Q_j^{(2)}) \otimes \cdots \otimes (\sum_{j=1}^{T} \eta_j Q_j^{(N)})\|_F^2$$

$$= \sum_{a=1}^{n_{Q^{(1)}}} \sum_{b=1}^{n_{Q^{(2)}}} \cdots \sum_{n=1}^{n_{Q^{(N)}}} \Big[\Big(\Sigma_{p=1}^{T} \sigma_{Q_p^{(1)}}(a) \cdot \sigma_{Q_p^{(2)}}(b) \cdots \sigma_{Q_p^{(N)}}(n)\Big) - \left(\sum_{p=1}^{T} \alpha_p \sigma_{Q_p^{(1)}}(a)\right) \Big(\Sigma_{p=1}^{T} \beta_p \sigma_{Q_p^{(2)}}(b)\Big) \cdots \Big(\Sigma_{p=1}^{T} \eta_p \sigma_{Q_p^{(N)}}(n)\Big)^2.$$

With this transformation, $\tilde{R}$ and its HOSVD are never needed.

- This is a nonlinear convex function of NT variables.

  As N and E $\uparrow$, practicality of problem transformation $\downarrow$.

  But, we can group automata!

Open Issues

- What values of N, T make transforming the preconditioning problem into a nonlinear optimization problem worthwhile?
- Is there a way to do a vector-$\tilde{R}$ multiplication without forming $\tilde{R}$ or the global Q? If so, we do not have to resort to the nonlinear optimization problem. We could use de Lathauwer's best rank-1 approximation algorithm to find the *optimal* $A, B, \ldots, N$ matrices.
- Does $(A \otimes B \otimes \cdots \otimes N \approx Q) \Rightarrow (A^{-1} \otimes B^{-1} \otimes \cdots \otimes N^{-1} \approx Q^{-1})$?

The End