Krylov subspace methods for computing stationary probability distributions of CTMCs

Olaf Schneider

Forschungszentrum Karlsruhe
Institut für Wissenschaftliches Rechnen

Joint GAMM–SIAM Conference on Applied Linear Algebra, 2006, Düsseldorf

Outline

Outline

Definition of a CTMC

- continuous-time stochastic process $X(t)$, $t \geq 0$, i. e. $\omega \mapsto X(t,\omega) = X(t)$ is a measurable function for each fixed $t \geq 0$,
- state space $\mathcal{S} = \{0, 1, 2, \ldots, n\}$ $\quad$ (n finite but large)
- Markov property: The probability of entering state j within the time period t depends only on the current state i and is independent from all other previous states.
- homogeneous in time:
  $\Pr\{X(s+t) = j \mid X(s) = i\}$ independet of s
- transition probababilities
  $p_{i,j}(t) = \Pr\{X(t) = j \mid X(0) = i\}$
- transition function (matrix) $P(t) = [p_{i,j}(t)]_{i,j\in\mathcal{S}} \in \mathbb{R}^{n\times n}$

State Space Distributions

- The state distribution $\boldsymbol{\pi}(t)$ of the chain is a vector, whose ith entry is the probability, that the process is in state i at the time t.
- $\boldsymbol{\pi}(t)\mathbf{1} = 1$ for all $t \geq 0$
- $\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)P(t)$

Fact

Transistion function $P(t)$ and initial distribution $\boldsymbol{\pi}(0)$ determine completely the CTMC.

Stationary (or invariant) distribution

$$\tilde{\boldsymbol{\pi}} = \tilde{\boldsymbol{\pi}}P(t)$$

Outline

Properties of the transition function $P(t)$

- Semigroup: $P(s+t) = P(s)P(t)$
- Identity: $P(0) = I$
- Enties are probabilities: $0 \leq p_{i,j}(t) \leq 1$
- Rows are state distributions: $P(t)\mathbf{1} = \mathbf{1}$
- Right continuous (standard): $\lim_{h\to 0} P(t+h) = P(t)$

Fact

All properties of P (and $\boldsymbol{\pi}$ and others) can be derived by using only these formulas.

Differentiability of $P(t)$

Main Identity

$$P(t) = e^{tQ}, \qquad t \geq 0$$

Infinisesimal Generator

$$Q = P'(0)$$

Kolmogorov Differential Equations

$$P'(t) = \frac{\mathrm{d}}{\mathrm{d}\,t} P(t) = QP(t) = P(t)Q$$

Properties of the infinitesimal generator Q

- $Q\mathbf{1} = \mathbf{0}$
- $Q = \lambda(\hat{P} - I)$, $\hat{P}$ stochastic
- Sign structure:

$$Q = \begin{pmatrix} - & + & \dots & + \\ + & - & \ddots & \vdots \\ \vdots & \ddots & \ddots & + \\ + & \dots & + & - \end{pmatrix}$$

Theorem

$-Q$ is a singular M-matrix with "property c".

Outline

1. Continuous Time Markov Chains
   - Definitions
   - Transition function and infinitesimal generator
   - Irreducible Markov Chains and Irreducible Matrices
   - Computing stationary distributions
2. Solving Singular Linear Equations
   - Abstract MR and OR projection methods
   - Regular and Singular Breakdowns
   - Krylov subspace methods

Definitions of irreducability

Definition

The Markov chain is irreducible if

$$\forall\, i,j \in \mathcal{S}, i \neq j : \exists\, t > 0 : \; p_{i,j}(t) > 0.$$

A matrix $Q \in \mathbb{C}^{n\times n}$, $n > 1$, is called *reducible* if there is some permutation matrix T and some integer r, $0 < r < n$ such that

$$T^\top Q T = \begin{bmatrix} A & B \\ O & C \end{bmatrix}$$

with $A \in \mathbb{C}^{r\times r}$, $B \in \mathbb{C}^{r\times(n-r)}$, $C \in \mathbb{C}^{(n-r)\times(n-r)}$ and a zero matrix $O \in \mathbb{C}^{(n-r)\times r}$. The matrix Q is called *irreducible* if it is not reducible.

Consequences of irreducability

Outline

Computing stationary distributions

Theorem (invariant vectors)

$$\tilde{\boldsymbol{\pi}} = \tilde{\boldsymbol{\pi}} P(t) \qquad \Longleftrightarrow \qquad \tilde{\boldsymbol{\pi}} Q = \mathbf{0}^\top$$

Reformulation of the problem:

- $A = -Q^\top$
- $x = \tilde{\boldsymbol{\pi}}^\top$
- $b = \mathbf{0}$

Singular linear equation

$$Ax = b$$

Computing stationary distributions

Theorem (invariant vectors)

$$\tilde{\boldsymbol{\pi}} = \tilde{\boldsymbol{\pi}} P(t), \quad \tilde{\boldsymbol{\pi}}\mathbf{1} = 1 \iff \qquad \tilde{\boldsymbol{\pi}} Q = \mathbf{0}^\top, \quad \tilde{\boldsymbol{\pi}}\mathbf{1} = 1$$

Reformulation of the problem:

- $A = -Q^\top$
- $x = \tilde{\boldsymbol{\pi}}^\top, \quad (x, \mathbf{1}) = 1$
- $b = \mathbf{0}$

Singular linear equation

$$Ax = b, \quad (x, \mathbf{1}) = 1$$

Computing stationary distributions

Theorem (invariant vectors)

$$\tilde{\boldsymbol{\pi}} = \tilde{\boldsymbol{\pi}} P(t), \quad \tilde{\boldsymbol{\pi}}\mathbf{1} = 1 \Longleftrightarrow \qquad \tilde{\boldsymbol{\pi}} Q = \mathbf{0}^\top, \quad \tilde{\boldsymbol{\pi}}\mathbf{1} = 1$$

Reformulation of the problem:

- $A = -Q^\top$
- $\boldsymbol{x} = \tilde{\boldsymbol{\pi}}^\top, \quad (\boldsymbol{x}, \mathbf{1}) = 1$
- $\boldsymbol{b} = \mathbf{0}$

Singular linear equation

$$A\boldsymbol{x} = \boldsymbol{b}, \quad (\boldsymbol{x}, \mathbf{1}) = 1$$

uniquely solvable for irreducible CTMC

Outline

Corrections from nested subspaces

Given

- initial guess x_0 such that

$$\mathbf{0} \neq r_0 = b - Ax_0$$

- nested *correction spaces*

$$\{\mathbf{0}\} = \mathcal{C}_0 \subset \mathcal{C}_1 \subset \mathcal{C}_2 \subset \cdots \subset \mathcal{C}_m \subset \mathcal{C}_{m+1} \subset \cdots$$

Iterative subspace correction method

The mth iterate for the solution of $Ax = b$ is

$$x_m = x_0 + c_m, \qquad c_m \in \mathcal{C}_m.$$

Reformulation into an approximation problem

- $\mathbf{r}_m := \mathbf{b} - A\mathbf{x}_m$ (residuals)
- $\mathbf{h}_m := A\mathbf{c}_m$
- $\mathbf{r}_m = \mathbf{r}_0 - \mathbf{h}_m$
- $\mathbf{h}_m$ is a suilable approximation of $\mathbf{r}_0$ from the mth *approximation space* $\mathcal{W}_m := A\mathcal{C}_m$

Reformulation into an approximation problem

- $\boldsymbol{r}_m := \boldsymbol{b} - A\boldsymbol{x}_m$ (residuals)
- $\boldsymbol{h}_m := A\boldsymbol{c}_m$
- $\boldsymbol{r}_m = \boldsymbol{r}_0 - \boldsymbol{h}_m$
- $\boldsymbol{h}_m$ is a suilable approximation of $\boldsymbol{r}_0$ from the mth *approximation space* $\mathcal{W}_m := A\mathcal{C}_m$
- nested sequence of approximation spaces

$$\{\mathbf{0}\} = \mathcal{W}_0 \subseteq \mathcal{W}_1 \subseteq \cdots \subseteq \mathcal{W}_{m-1} \subseteq \mathcal{W}_m \subseteq \cdots$$

Reformulation into an approximation problem

- $\boldsymbol{r}_m := \boldsymbol{b} - A\boldsymbol{x}_m$ (residuals)
- $\boldsymbol{h}_m := A\boldsymbol{c}_m$
- $\boldsymbol{r}_m = \boldsymbol{r}_0 - \boldsymbol{h}_m$
- $\boldsymbol{h}_m$ is a suilable approximation of $\boldsymbol{r}_0$ from the mth *approximation space* $\mathcal{W}_m := A\mathcal{C}_m$
- nested sequence of approximation spaces

$$\{\mathbf{0}\} = \mathcal{W}_0 \subseteq \mathcal{W}_1 \subseteq \cdots \subseteq \mathcal{W}_{m-1} \subseteq \mathcal{W}_m \subseteq \cdots$$

- given $\boldsymbol{h}_m$ the correction $\boldsymbol{c}_m$ is determined only up to an arbitrary component from $\mathcal{C}_m \cap \mathcal{N}(A)$

MR and OR Approximations

The MR approximation $\boldsymbol{h}_m^{\mathrm{MR}}$ is defined as the best approximation to $\boldsymbol{r}_0$ from $\mathcal{W}_m$, i. e.

$$\boldsymbol{h}_m^{\mathrm{MR}} \in \mathcal{W}_m, \qquad \boldsymbol{r}_m^{\mathrm{MR}} := \boldsymbol{r}_0 - \boldsymbol{h}_m^{\mathrm{MR}} \perp \mathcal{W}_m.$$

The related correction $\boldsymbol{c}_m^{\mathrm{MR}}$ with $\boldsymbol{h}_m^{\mathrm{MR}} = A\boldsymbol{c}_m^{\mathrm{MR}}$ and $\boldsymbol{x}_m^{\mathrm{MR}} = \boldsymbol{x}_0 + \boldsymbol{c}_m^{\mathrm{MR}}$ solves $\|\boldsymbol{b} - A\boldsymbol{x}_m^{\mathrm{MR}}\| = \min_{\boldsymbol{c}\in\mathcal{C}_m}\|\boldsymbol{r}_0 - A\boldsymbol{c}\|$.

Given an auxiliary test space $\mathcal{V}_m$ the OR approximation $\boldsymbol{h}_m^{\mathrm{OR}}$ is obtained by imposing

$$\boldsymbol{h}_m^{\mathrm{OR}} \in \mathcal{W}_m, \qquad \boldsymbol{r}_m^{\mathrm{OR}} := \boldsymbol{r}_0 - \boldsymbol{h}_m^{\mathrm{OR}} \perp \mathcal{V}_m.$$

Usual choise $\mathcal{V}_m := \mathrm{span}\{\boldsymbol{r}_0\} + \mathcal{W}_{m-1}$, called *residual space* since it contains $\boldsymbol{r}_m = \boldsymbol{r}_0 - \boldsymbol{h}_m$, $\boldsymbol{h}_m \in \mathcal{W}_m$.

MR methods unsing an orthonormal basis of $\mathcal{V}_m$

- $\boldsymbol{v}_1 = \frac{1}{\beta}\boldsymbol{r}_0$
- For $m = 1, 2, \ldots$

$$\boldsymbol{v}_1 = \boldsymbol{r}_0/\beta, \text{ where } \beta := \|\boldsymbol{r}_0\|,$$

$$\boldsymbol{v}_{m+1} = \frac{(I - P_{\mathcal{V}_m})A\boldsymbol{c}_m}{\|(I - P_{\mathcal{V}_m})A\boldsymbol{c}_m\|},$$

where $\boldsymbol{c}_m \in \mathcal{C}_m \setminus \mathcal{C}_{m-1}$

- breakdown if $A\boldsymbol{c}_m \in \mathcal{V}_m$
- denote the breakdown index by L

Arnoldi-type decomposition

$$AC_m = V_{m+1}\tilde{H}_m = V_m H_m + \begin{bmatrix}\mathbf{0} & \dots & \mathbf{0} & \eta_{m+1,m}\boldsymbol{v}_{m+1}\end{bmatrix},$$

where $C_m := \begin{bmatrix}\boldsymbol{c}_1 & \cdots & \boldsymbol{c}_m\end{bmatrix}$ and $V_{m+1} := \begin{bmatrix}\boldsymbol{v}_1 & \cdots & \boldsymbol{v}_{m+1}\end{bmatrix}$
Reduced the least squares problem in the coordinate space:

$$\begin{aligned}\|\boldsymbol{b} - A\boldsymbol{x}_m^{\mathrm{MR}}\| &= \|\boldsymbol{r}_0 - AC_m\boldsymbol{y}_m^{\mathrm{MR}}\| \\ &= \|V_{m+1}(\beta\boldsymbol{u}_1^{(m+1)} - \tilde{H}_m\boldsymbol{y}_m^{\mathrm{MR}})\| =\end{aligned}$$

$$\|\beta\boldsymbol{u}_1^{(m+1)} - \tilde{H}_m\boldsymbol{y}_m^{\mathrm{MR}}\|_2 = \min_{\boldsymbol{y}\in\mathbb{C}^m}\|\beta\boldsymbol{u}_1^{(m+1)} - \tilde{H}_m\boldsymbol{y}\|_2$$

Outline

1. Continuous Time Markov Chains
   - Definitions
   - Transition function and infinitesimal generator
   - Irreducible Markov Chains and Irreducible Matrices
   - Computing stationary distributions
2. Solving Singular Linear Equations
   - Abstract MR and OR projection methods
   - Regular and Singular Breakdowns
   - Krylov subspace methods

Regular breakdown
of an MR method in step m

Quite the same as for non-singular matrix A:

- $A\boldsymbol{c}_m \in \mathcal{V}_m$ and H_m is regular
- $\boldsymbol{r}_0 \in \mathcal{W}_m = \operatorname{span}\{A\boldsymbol{c}_1, \ldots, A\boldsymbol{c}_m\}$
- $\boldsymbol{x}_m^{\mathrm{MR}} = \boldsymbol{x}_0 + \beta C_m H_m^{-1} \boldsymbol{u}_1^{(m)}$
  is a proper solution of the original linear equation

- $\mathcal{W}_m = \mathcal{V}_m$

(these conditions are equivalent).

Regular breakdown

of an MR method in step m

Quite the same as for non-singular matrix A:

- $A\boldsymbol{c}_m \in \mathcal{V}_m$ and H_m is regular
- $\boldsymbol{r}_0 \in \mathcal{W}_m = \text{span}\{A\boldsymbol{c}_1, \ldots, A\boldsymbol{c}_m\}$
- $\boldsymbol{x}_m^{\text{MR}} = \boldsymbol{x}_0 + \beta C_m H_m^{-1} \boldsymbol{u}_1^{(m)}$
  is a proper solution of the original linear equation,
  i. e. $A\boldsymbol{x} = \boldsymbol{b}$ is necessary consistent.
- $\mathcal{W}_m = \mathcal{V}_m$

(these conditions are equivalent).

Fact

If A is regular, the method breaks down regularly.

Singular breakdown

of an MR method in step m

Several equivalent characterizations:

- no unique solution of the least squares problem
- rank deficiency of the Hessenberg matrix
- R_m (and $\tilde{H}_m^{\mathsf{H}} \tilde{H}_m$) singular
- $\mathcal{W}_m = \mathcal{W}_{m-1}$
- $\mathcal{C}_m \cap \mathcal{N}(A) \neq \{\mathbf{0}\}$

Singular breakdown

of an MR method in step m

Several equivalent characterizations:

- no unique solution of the least squares problem
- rank deficiency of the Hessenberg matrix
- R_m (and $\tilde{H}_m^{\mathsf{H}}\tilde{H}_m$) singular
- $\mathcal{W}_m = \mathcal{W}_{m-1}$
- $\mathcal{C}_m \cap \mathcal{N}(A) \neq \{\mathbf{0}\}$

Fact

If the linear system is inconsistent, every breakdown is a singular one.

Outline

Corrections from Krylov spaces

Breakdown of Krylov subspace methods

Using Krylov subspace methods for Markov problems

Supporting facts:

- The linear system is consistent.
- $ind A = 1$
- $\mathbf{x}_L^{MR} = \mathbf{x}_L^{OR} = (I - A^{\#}A)\mathbf{x}_0$
- choise $\mathbf{x}_0 = \frac{1}{n}\mathbf{1}$ ensures $\mathbf{x}_0 \notin \mathcal{R}(A)$
- If the chain is irreducible then $\dim \mathcal{N}(A) = 1$, i. e. there is a unique stationary solution.

Summary

- Application of Krylov methods to singular problems is not a good idea in general.
- MR Krylov methods are well suited for computing stationary distributions of CTMCs.

- Open Questions
  - What about GCR?
  - How to determ all stationary vectors of an reducilbe chain?

For Further Reading I

A. Author.
Handbook of Everything.
Some Press, 1990.

S. Someone.
On this and that.
Journal of This and That, 2(1):50–100, 2000.