Iterative component-wise bounds for the steady-state distribution of a Markov chain

J.M. Fourneau

Laboratoire PRiSM, CNRS UMR 8144

Université de Versailles St-Quentin

joint work with Ana Bušić (INRIA)

Finite DTMC P with stationary solution

- Computation of the steady-state distribution π_P
- CTMC (same method after uniformization)
- Proved convergence (i.e. one can build S such that $\pi_P \in S$ and S as small as needed).
- Compute bounds rather than π_P

Changing the point of view

- We consider non negative matrices (not stochastic)
- We use the element wise comparison of matrices and vectors ($\preceq_{el}$)
- Dynamical systems (i.e. non stochastic) based on (max,+) or (min,+) sequences
- Convergence to π_P or bounds of π_P
- Depending on some quantities easily computed on P

Outline of the paper

- Definition of ∇_P.
- Algorithms and Results when $\nabla_P \neq \vec{0}$ (the easy case).
- Algorithms and Results when $\nabla_P = \vec{0}$ (the hard case).
- Aggregation and Bounds (a new solution).

An old and simple Idea

- **Lemma 1** *As* $\pi_P = \pi_P P$*, and as* $\pi(j)$ *is between* 0 *and* 1 *for all* j*, then we have:*

$$Min_i P(i,j) \leq \pi_P(j) = \sum_i \pi_P(i) P(i,j) \leq Max_i P(i,j)$$

- Definition: $\nabla_P(j) = Min_i P(i,j)$ and $\Delta_P(j) = Max_i P(i,j)$

$\nabla_P \neq \vec{0}$, Lower Bound

Algorithm Iterate ∇ Lower Bound (I∇L)

- Parameters $d1 \preceq_{el} \pi_P$, $d2 \preceq_{el} \nabla_P$ and $d2 \neq \vec{0}$.
- Initialization $x^{(0)} = d1$
- Iteration $x^{(k+1)} = max(x^{(k)}, x^{(k)}P + d2(1 - ||x^{(k)}||_1))$

Theorem 1 *Let P be an irreducible stochastic matrix. Assume that the steady-state probability distribution π_P exists. If $\nabla_P \neq \vec{0}$, Algorithm (I∇L) provides lower bounds for all components of π_P and converges to π_P for any value of the parameters $d1$ and $d2$.*

Exemple 1

Example 1 $P1 = \begin{pmatrix} 0.5 & 0.2 & 0.0 & 0.3 \\ 0.0 & 0.2 & 0.5 & 0.3 \\ 0.1 & 0.2 & 0.5 & 0.2 \\ 0.0 & 0.6 & 0.0 & 0.4 \end{pmatrix}$. $\nabla_{P1} = (0.0, 0.2, 0.0, 0.2)$.

Algorithm ($I\nabla L$) with $d1 = d2 = \nabla_{P1}$ gives the following sequence of lower bounds for the probabilities.

k	*1*	*2*	*3*	*4*	*Residual*
1	*0.00000*	*0.28000*	*0.10000*	*0.26000*	*0.36000*
5	*0.04630*	*0.31643*	*0.29723*	*0.29338*	*0.04665*
10	*0.06236*	*0.31886*	*0.31770*	*0.29743*	*0.00362*
15	*0.06371*	*0.31912*	*0.31903*	*0.29783*	*0.00028*
20	*0.06382*	*0.31914*	*0.31914*	*0.29787*	*0.00002*

$\nabla_P \neq \vec{0}$, Upper Bound

——— Algorithm Iterate ∇ Upper Bound (I∇U) ———

- Parameters $d3 \succeq_{el} \pi_P$, $d2 \preceq_{el} \nabla_P$ and $d2 \neq \vec{0}$.
- Initialize: $y^{(0)} = d3$.
- Iterate: $y^{(k+1)} = min(y^{(k)}, y^{(k)}P + d2(1 - ||y^{(k)}||_1))$.

Theorem 2 *Let P be an irreducible stochastic matrix. Assume that the steady-state probability distribution π_P exists. If $\nabla_P \neq \vec{0}$, Algorithm I∇U provides a sequence of non increasing upper bounds for all the components of π_P and leads to π_P.*

Convergence

- At every step k, $x^{(k)} \preceq_{el} \pi_P \preceq_{el} y^{(k)}$
- And $x^{(k)}$ is increasing, while $y^{(k)}$ is decreasing. $||y^{(k)} - x^{(k)}||_1$ converges to 0.
- **Example 2** *Again* $P1$ *(Example 1), after* 20 *iterations, combining both algorithms we have the following intervals for the steady-state distribution:*

1	*(0.06382,0.06384)*
2	*(0.31914,0.31915)*
3	*(0.31914 , 0.31916)*
4	*(0.29787,0.29787)*

Complexity

- nz non zero entries, n size of the state space
- Computing ∇: $\theta(nz)$,
- Per Iteration: Vector Matrix multiplication $\theta(nz)$ for a sparse matrix. + Linear complexity operations.
- See also: Kronecker representation.
- Number of iterations: convergence upper bounded by a geometric with rate $||\nabla_P||_1$
- Speed of convergence: easily obtained from the matrix

Simplification of Matrices

- Assumes that computing P is too difficult...
- Typically censoring, ideal aggregation, stochastic complement....
- But you know how to compute Q an element wise lower bound of P. Assume that $\nabla_Q \neq \vec{0}$.
- **Property 1** *Use Q instead of P in $I\nabla L$ to obtain an increasing sequence upper bounded by π_P.*
- **Property 2** *Similarly, if $P \preceq_{el} R$, use R instead of P in $I\nabla U$ to obtain (under some technical constraints) a decreasing sequence lower bounded by π_P.*

The hard case: $\nabla_P = \vec{0}$, Lower bound

————Simple Lower Bound Algorithm—-(SLB)——

- Initialization: $x^{(0)} = d \preceq_{el} \pi_P$.
- Iteration: $x^{(k+1)} = max(x^{(k)}, x^{(k)}P)$.

———————————————————-

Theorem 3 *For d such that $0 \preceq_{el} d \preceq_{el} \pi_P$, SLB Algorithm converges. Moreover, if $d \neq \vec{0}$, then the limit vector x of the sequence $x^{(k)}$, $k \geq 0$ is a multiple of π_P (i.e. $x = \frac{\pi_P}{||x||_1}$) and $||x||_1 \leq 1$.*

The hard case: $\nabla_P = \vec{0}$, Upper bound

———Simple Upper Bound Algorithm—(SUB)———

- Initialization: $z^{(0)} = d3 \succeq_{el} \pi_P$.
- Iteration: $z^{(k+1)} = min(z^{(k)}, z^{(k)}P)$.

Remark 1 *Again one can use* $d3 = \Delta_P$ *in the initialization step of this algorithm.*

Theorem 4 *Algorithm SUB gives a decreasing sequence of upper bound vector for* π_P *which converges to a multiple of* π_P*.*

Problems

- How to find $d \preceq_{el} \pi_P$ and $d \neq \vec{0}$
- A solution is in the paper based on stochastic complement
- Test for convergence: difference between successive values of $z^{(k)}$ (same problem from $x^{(k)}$)
- Provides two approximations: $\frac{z^{(k)}}{||z^{(k)}||_1}$ and $\frac{x^{(k)}}{||x^{(k)}||_1}$.
- How to build an accurate bound from bounds of a multiple of π_P?

Exemple

Example 3 $P7 = \begin{pmatrix} 0.2 & 0.3 & 0.2 & 0.3 \\ 0.5 & 0.1 & 0.0 & 0.4 \\ 0.0 & 0.6 & 0.4 & 0.0 \\ 0.5 & 0.0 & 0.2 & 0.3 \end{pmatrix}$. $\nabla_{P7} = (0.0, 0.0, 0.0, 0.0)$ *and* $\Delta_{P7} = (0.5, 0.6, 0.4, 0.4)$. *Assume that we have found a lower bound of* π_{P7} *which is equal to* $(0.0, 0.0, 0.15, 0.0)$. *The algorithms SLB and SUB provide the following lower (first table) and upper (second table) bounds:*

k	*1*	*2*	*3*	*4*	*Residual*
1	*0.000000*	*0.090000*	*0.150000*	*0.000000*	*0.760000*
10	*0.195324*	*0.161590*	*0.150000*	*0.166461*	*0.326625*
20	*0.234766*	*0.177670*	*0.150000*	*0.200543*	*0.237021*
40	*0.242429*	*0.180793*	*0.150000*	*0.207164*	*0.219614*
60	*0.242641*	*0.180880*	*0.150000*	*0.207348*	*0.219131*

k	*1*	*2*	*3*	*4*	*Residual*
1	*0.500000*	*0.450000*	*0.340000*	*0.400000*	*-0.690000*
10	*0.471310*	*0.352347*	*0.291323*	*0.400000*	*-0.514980*
20	*0.468221*	*0.349080*	*0.289445*	*0.400000*	*-0.506746*
40	*0.468085*	*0.348936*	*0.289362*	*0.400000*	*-0.506384*
60	*0.468085*	*0.348936*	*0.289362*	*0.400000*	*-0.506383*

Improvements

- Gauss-Seidel effect
- Joined computation of lower and upper bound based on:

Lemma 2 *For* $x^{(k)} \preceq_{el} \pi_P \preceq_{el} z^{(k)}$*, we have:*

1. $x^{(k)}P + \nabla(1 - ||x^{(k)}||_1) \preceq_{el} \pi_P \preceq_{el} z^{(k)}P + \nabla(1 - ||z^{(k)}||_1)$
2. *Similarly,*
   $z^{(k)}P + \Delta(1 - ||z^{(k)}||_1) \preceq_{el} \pi_P \preceq_{el} x^{(k)}P + \Delta(1 - ||x^{(k)}||_1)$

Aggregation

- Assume that P is finite, irreducible, and $\nabla_P = \vec{0}$.
- Assume that P is not lumpable.
- Assume that we have found a partition $\mathcal{A}_1, \ldots, \mathcal{A}_m$ $(m \geq 2)$ of the state-space such it exists a non empty subset $\mathcal{A}_k$ such that for all state j in the state space $P(j, \mathcal{A}_k) > 0$.

Ideal Aggregates

- $R = WPV$, V is a collector matrix, W a distributor matrix
- $V(i, k) = 1$ if state i is in subset $\mathcal{A}_k$
- Columns of W contains the conditional distribution of states for a subset
- We must know π_P to compute the aggregated matrix W and R
- If the aggregated Markov chain is an ideal aggregation of P, then $\pi_R = \pi_P V$.

Bounds on Matrices

- Entry (i,k) of matrix PV is $P(i,\mathcal{A}_k)$.
- Entry (k,j) of matrix W is the conditional distribution if state i is in subset $\mathcal{A}_k$ and 0 otherwise.
- Entry (l,k) of matrix WPV is the convex sum of $P(i,\mathcal{A}_k)$ for $i \in \mathcal{A}_l$ and for all distributor matrix W we have:

$$Min_{i\in\mathcal{A}_l} P(i,\mathcal{A}_k) \leq (WPV)(l,k) \leq Max_{i\in\mathcal{A}_l} P(i,\mathcal{A}_k)$$

- Finally, $L \preceq_{el} WPV \preceq_{el} U$
  where $L(l,k) = Min_{i\in\mathcal{A}_l} P(i,\mathcal{A}_k)$

Bounds on steady-state distribution

- $L \preceq_{el} WPV \preceq_{el} U$
- Because of the partition $\nabla_R \neq \vec{0}$,
- Apply Algorithm $I\nabla L$ on L (a lower bound of R) to obtain an increasing sequence upper bounded by $\pi_P V$.

Conclusion

- New computation scheme for the steady-state distribution of DTMC
- Compatible with tensor and sparse representation
- Provide bounds at every step
- Proved convergence if $\nabla \neq \vec{0}$
- Simplification of matrices implies component-wise bounds
- Mixed approach with 'Ideal Aggregates' to obtain component-wise bounds for the aggregates.
- Works well with Google matrix