The Physics of Matrix-Analytic Algorithms.

Peter Taylor

The University of Melbourne

November 16, 2015

Discrete-time Markov chains (DTMCs)

These are named after Andrei Andreevich Markov (1856-1922).

- Secondary education: good in mathematics but performed poorly in other subjects.
- 1874: PhD, University of Petersburg (under Chebyshev).
- 1878: gold medal for his scientific work.
- Research areas: number theory, the approximation of functions, the problem of moments, the calculus of finite differences.
- Best remembered for: a theory of chains of stochastic processes (Markov chains).

Markov is my academic great-great-grandfather. See the Mathematics Genealogy Project (http://genealogy.math.ndsu.nodak.edu/search.php).

Discrete-time Markov chains

A random sequence $\{X_n, n \geq 0\}$ with a countable state space (for example, $\{0, 1, 2, \cdots\}$) forms a DTMC if

$$P(X_{n+1} = k | X_n = j, X_{n-1} = x_{n-1}, \cdots, X_0 = x_0) = P(X_{n+1} = k | X_n = j).$$

This enables us to write

$$P(X_{n+1} = k | X_n = j) = p_{jk}(n).$$

Furthermore, we commonly assume that the transition probabilities $p_{jk}(n)$ do not depend on n, in which case the DTMC is called homogeneous and we write $p_{jk}(n) = p_{jk}$.

Discrete-time Markov chains

For a homogeneous DTMC, the transition matrix is the matrix with rows and columns corresponding to the states of the process and whose jkth entry is p_{jk}. So

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{pmatrix}.$$

Discrete-time Markov chains

The n-step transition probabilities $P(X_{m+n} = j | X_m = i)$ of a homogeneous DTMC do not depend on m. For $n = 1, 2, \cdots$, we denote them by

$$p_{ij}^{(n)} = P(X_{m+n} = j | X_m = i).$$

It is also convenient to use the notation

$$p_{ij}^{(0)} := \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i. \end{cases}$$

Discrete-time Markov chains

The Chapman-Kolmogorov equations show how we can calculate the $p_{ij}^{(n)}$ from the p_{ij}.

For $n = 1, 2, \cdots$ and any $r = 1, 2, \cdots, n$,

$$p_{ij}^{(n)} = \sum_k p_{ik}^{(r)} p_{kj}^{(n-r)}.$$

Discrete-time Markov chains

If we define the n-step transition matrix as

$$P^{(n)} = \begin{pmatrix} p_{11}^{(n)} & p_{12}^{(n)} & \ddots & \ddots \\ p_{21}^{(n)} & p_{22}^{(n)} & p_{23}^{(n)} & \ddots \\ \ddots & \ddots & \ddots & \ddots \end{pmatrix},$$

then the Chapman-Kolmogorov equations can be written in the matrix form

$$P^{(n)} = P^{(r)} P^{(n-r)}$$

with $P^{(1)} = P$. By mathematical induction, it follows that

$$P^{(n)} = P^n,$$

the nth power of P.

Discrete-time Markov chains

If we have

- an initial distribution $\boldsymbol{\pi}^0 = (\pi_1^0, \ldots, \pi_n^0)$, where $\pi_j^0 = P(X_0 = j)$, for all j, and
- the transition matrix P,

we can (in principle) use the Markov property to derive the finite dimensional distributions.

For $k \geq 1$ and $n_1 < \cdots < n_k \in Z_+$,

$$\begin{aligned} & P(X_{n_0} = x_0, X_{n_1} = x_1, X_{n_2} = x_2, \cdots, X_{n_k} = x_k) \\ = \quad & \boldsymbol{\pi^0}_{x_0} P^{n_1}_{x_0 x_1} [P^{n_2 - n_1}]_{x_1 x_2}, \ldots, [P^{n_k - n_{k-1}}]_{x_k x_{k-1}}. \end{aligned}$$

Discrete-time Markov chains

We are often interested in the stationary distribution of a DTMC. This can be interpreted as giving the long-run proportion of time that the chain spends in each of its states.

The stationary distribution does not always exist. However if it does exist, then it is the unique solution to the system of linear equations

$$\pi P = \pi,$$

with $\sum_j \pi_j = 1$.

We often test whether a DTMC has a stationary distribution by attempting to solve these equations.

Continuous-time Markov chains (CTMCs)

A non-negative integer valued stochastic process $\{X_t : \ t \geq 0\}$ in continuous time is said to be a Continuous-Time Markov Chain if, for all $k \geq 1$, $t_1 < t_2 < \cdots < t_{k+1}$ and non-negative integers $i_1, i_2, \ldots, i_{k+1}$,

$$\begin{aligned} & P(X_{t_{k+1}} = i_{k+1} | X_{t_1} = i_1, \cdots, X_{t_k} = i_k) \\ = \quad & P(X_{t_{k+1}} = i_{k+1} | X_{t_k} = i_k). \end{aligned}$$

If $P(X_{t+h} = k | X_h = j) = P(X_t = k | X_0 = j) \equiv p_{jk}^{(t)}$ does not depend on h, we say the CTMC is homogeneous.

Continuous-time Markov chains

Observe that

$$\begin{aligned} p_{ij}^{(s+t)} &= \sum_k P(X_{s+t} = j | X_s = k, X_0 = i) P(X_s = k | X_0 = i) \\ &= \sum_k p_{ik}^{(s)} p_{kj}^{(t)}. \end{aligned}$$

These are the Chapman-Kolomogorov equations for a CTMC. In matrix form, we write $P^{(t)} = (p_{jk}^{(t)})$. Then, for $s, t \geq 0$, the Chapman-Kolmogorov equations can be expressed in the form

$$P^{(t+s)} = P^{(t)} P^{(s)}.$$

We can show that $P^{(t)}$ is stochastically continuous in the sense that $P^{(t+h)} \to P^{(t)}$ as $h \to 0$.

Continuous-time Markov chains

If $t = m$ (a positive integer), the CChapman-Kolmogorov equations tell us that $P^{(m)} = (P^{(1)})^m$ and our hope is fulfilled. If t and h are nonnegative real numbers, we can write

$$\begin{aligned} \frac{P^{(t+h)} - P^{(t)}}{h} &= P^{(t)} \left[\frac{P^{(h)} - I}{h} \right] \\ &= \left[\frac{P^{(h)} - I}{h} \right] P^{(t)} \end{aligned}$$

This suggests that we should investigate the existence of the derivative

$$Q \equiv \lim_{h \to 0^+} \frac{P^{h)} - I}{h}.$$

Continuous-time Markov chains

The matrix Q is called the generator of the CTMC. In practical modelling with CTMCs it is taken as the data of the model.

Writing $Q = (q_{jk})$, for $j \neq k$, $q_{jk} \geq 0$ is the transition rate from state j to state k and $q_{jj} \leq 0$ is such that $-q_{jj}$ is the total transition rate leaving state j. We usually take

$$\sum_k q_{jk} = 0$$

for all j.

Continuous-time Markov chains

For non-explosive CTMCs, the matrix Q determines the transition probability completely by solving the Kolmogorov backward or forward equations to get

$$\begin{aligned} P^{(t)} &= \exp(tQ) \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} t^k Q^k, \end{aligned}$$

subject to $P^{(0)} = I$.

The stationary distribution π satisfies

$$\pi Q = \mathbf{0},$$

with $\sum_j \pi_j = 1$.

Continuous-time Markov chains

Starting at an initial state $X_0 = j$, a CTMC stays in j for an exponentially-distributed time with parameter $-a_{jj}$.

Then it jumps to a state $k \neq j$ with probability $-a_{jk}/a_{jj}$ and stays there for a random time which is exponentially-distributed with parameter $-a_{kk}$ and independent of anything that has happened previously, and then it jumps to ℓ, and so on ...

In mathematical biology, this simple observation is often called Gillespie's Algorithm, because Gillespie discovered it for himself in 1976. See

Gillespie D.T., A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions, Journal of Computational Physics 22, 403–434.

Matrix Analytic Models

- In the 1970s and 1980s Marcel Neuts proposed a class of techniques for analysing Markov chains with block-structured transition matrices that have become known as matrix-analytic methods.
- More recently, there has been interest in general Markov additive models, which can be thought of as Markov-modulated Levy processes.
- The interaction of mathematical analysis and physical insight has played an important role in the development of results in this area.
- There is an emphasis on computability of performance measures and, in particular, on algorithmic development.

Block-structured Markov chains

A discrete-time *Markov chain of GI/M/1-type* has a two-dimensional state space. The first dimension is countably-infinite and the second dimension is finite. When the chain is in state (k, i), we say that it is in *level* k and *phase* i. With a suitable ordering of the states, the transition matrix can be written in the form

$$P = \begin{bmatrix} \tilde{A}_1 & A_0 & 0 & 0 & \cdots \\ \tilde{A}_2 & A_1 & A_0 & 0 & \cdots \\ \tilde{A}_3 & A_2 & A_1 & A_0 & \cdots \\ \tilde{A}_4 & A_3 & A_2 & A_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Block-structured Markov chains

The terminology *Markov chain of GI/M/1-type* comes from the fact that the embedded Markov chain generated by a $GI/M/1$ queue observed at arrival points has a transition matrix of the form

$$P_G = \begin{bmatrix} \tilde{a}_1 & a_0 & 0 & 0 & \cdots \\ \tilde{a}_2 & a_1 & a_0 & 0 & \cdots \\ \tilde{a}_3 & a_2 & a_1 & a_0 & \cdots \\ \tilde{a}_4 & a_3 & a_2 & a_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

where a_k is the probability that there are k services during an inter-arrival interval and $\tilde{a}_k = \sum_{\ell=k}^{\infty} a_\ell$.

Block-structured Markov chains

A discrete-time *Markov chain of $M/G/1$-type* has a state space of identical structure. Rather than being block skip-free to the right, it is block skip-free to the left, so that its transition matrix can be written in the form

$$P = \begin{bmatrix} \tilde{A}_1 & \tilde{A}_2 & \tilde{A}_3 & \tilde{A}_4 & \cdots \\ A_0 & A_1 & A_2 & A_3 & \cdots \\ 0 & A_0 & A_1 & A_2 & \cdots \\ 0 & 0 & A_0 & A_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Block-structured Markov chains

Its name comes from the fact that the embedded Markov chain generated by a $M/G/1$ queue observed at departure points has a transition matrix of the form

$$P = \begin{bmatrix} \tilde{a}_1 & \tilde{a}_2 & \tilde{a}_3 & \tilde{a}_4 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & a_0 & a_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

where a_k is the probability that there are k arrivals during a service time and $\tilde{a}_k$ is the probability that the first service time in a busy period will finish with $k-1$ customers waiting in the queue. In this simple case, $\tilde{a}_k = a_{k-1}$.

Block-structured Markov chains

From a modelling point of view, the second dimension has many uses. For example, it can be used to denote

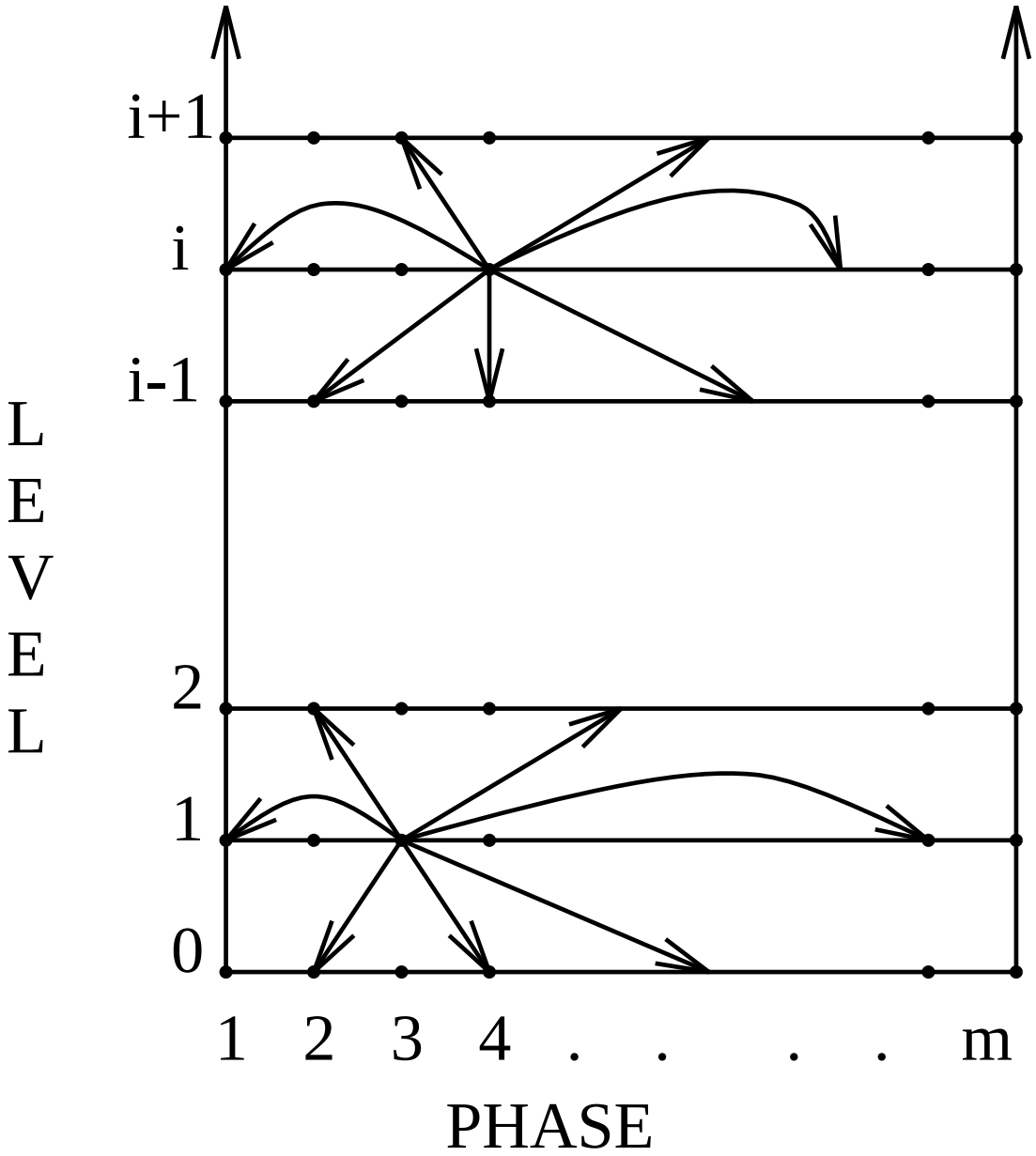
- the state of an independently-moving environment,
- a number of transmitting sources,
- the progress of one or more phase-type random variables,
- the number of individuals in an interacting species,
- the underlying state of a hidden Markov chain model,
- etc.

Block-structured Markov chains

Markov chains that are both of $GI/M/1$-type and $M/G/1$-type are known as *Quasi-Birth-and-Death Processes (QBDs)*. Their transition matrices can be written in the form

$$P = \begin{bmatrix} \tilde{A}_1 & \tilde{A}_0 & 0 & 0 & \cdots \\ A_2 & A_1 & A_0 & 0 & \cdots \\ 0 & A_2 & A_1 & A_0 & \cdots \\ 0 & 0 & A_2 & A_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

A Quasi-Birth-and-Death Process



Chains of $GI/M/1$-type

The matrix $A \equiv \sum_{k=0}^{\infty} A_k$ describes transitions in the phase, independently of the level.

For a discrete-time chain of $GI/M/1$-type, let $\boldsymbol{x}$ be the solution to

$$\boldsymbol{x}A = \boldsymbol{x}.$$

Then the chain is positive recurrent, null recurrent or transient according as

$$\boldsymbol{x}A_0\boldsymbol{e}' - \boldsymbol{x}\left[\sum_{k=2}^{\infty}(k-1)A_k\right]\boldsymbol{e}',$$

is less than, equal to or greater than zero.

Chains of $GI/M/1$-type

Write the stationary distribution of a discrete-time positive recurrent chain of $GI/M/1$-type as $\pi = (\pi_0, \pi_1, \ldots)$. Then there exists a matrix R such that

$$\pi_n = \pi_0 R^n.$$

The vector π_0 satisfies

$$\pi_0 \left[\sum_{k=0}^{\infty} R^k \tilde{A}_{k+1} \right] = \pi_0.$$

This is the well-known *matrix-geometric form* of the stationary distribution.

Chains of $GI/M/1$-type

The matrix R is the minimal nonnegative solution to the matrix equation

$$\sum_{k=0}^{\infty} R^k A_k = R.$$

The (i,j)th entry of the matrix R is the expected number of visits to phase j of level 1 before first return to level 0 conditional on the process starting in phase i of level 0.

In general, R has spectral radius which is less than or equal to one, and the chain is positive recurrent if and only if the spectral radius of R is less than one.

Chains of $M/G/1$-type

To derive the stationary distribution of chains of $M/G/1$-type, we use the fact that $\pi = (\pi_0, \pi_1, \ldots, \pi_n)$ is proportional to the stationary distribution of the finite-state Markov chain with transition matrix

$$P = \begin{bmatrix} \tilde{A}_1 & \tilde{A}_2 & \tilde{A}_3 & \cdots & \tilde{A}_n & \sum_{k=0}^{\infty} \tilde{A}_{k+n+1} G^k \\ A_0 & A_1 & A_2 & A_3 & \cdots & \vdots \\ 0 & A_0 & A_1 & A_2 & \cdots & \vdots \\ 0 & 0 & A_0 & A_1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \sum_{k=0}^{\infty} A_{k+2} G^k \\ 0 & 0 & 0 & \cdots & A_0 & \sum_{k=0}^{\infty} A_{k+1} G^k \end{bmatrix}.$$

Chains of $M/G/1$-type

The (i,j)th entry of the matrix G is the probability that the chain hits level $k-1$ in finite time, and does so in phase j, given that it starts in phase i of level k.

Clearly G is substochastic and it is stochastic if and only if the chain is recurrent.

Elementary arguments show that the matrix G is the minimal nonnegative solution to the matrix equation

$$\sum_{k=0}^{\infty} A_k G^k = G.$$

Chains of $M/G/1$-type

G is a matrix of probabilities, rather than a matrix of expected values (as R is). As such, it is a 'nicer' object to work with. Furthermore, for a QBD, the matrix R can be written in terms of the matrix G via the relation

$$R = A_2 \left[I - A_1 - A_0 G \right]^{-1}$$

and, for a chain of $GI/M/1$-type, the matrix R can be written in terms of the matrix G for the *dual chain* of $M/G/1$-type. For this reason, we concentrate on algorithms for calculating G.

To keep the notation simple, I shall discuss the QBD special case.

Calculating the matrix G

In the QBD special case, G is the minimal nonnegative solution to the matrix quadratic equation

$$A_2 + A_1 G + A_0 G^2 = G.$$

This equation has an analytic solution only in a few special cases.

In general, we have to resort to numerical solution.

A simple procedure

For an irreducible QBD, A_1 is invertible. So, an obvious first approach to solving this equation is to transform it into a fixed-point equation:

$$\begin{aligned}(I - A_1)G &= A_2 + A_0 G^2 \\ \Rightarrow G &= (I - A_1)^{-1}\left[A_2 + A_0 G^2\right]\end{aligned}$$

and use the iterative procedure

$$G_{n+1} = (I - A_1)^{-1}\left[A_2 + A_0 G_n^2\right]$$

with $G_0 = 0$.

A simple procedure

Neuts showed that, with this iteration, G_n does converge to G.

Furthermore, except when the QBD is null-recurrent, this convergence is *linear*.

That is, there exists a constant $\gamma \in (0, 1)$ such that

$$\limsup_{n \to \infty} ||G_n - G||^{1/n} = \gamma.$$

Physical interpretations

The type of question that we shall be interested in is

Can we give a physical interpretation to the nth iterate of procedures such as the one described above?

For Neuts' original iteration mentioned above, this question has not had a precise answer until about ten year's ago. It can be understood in terms of iterations for tree-structured QBDs.

In general, to understand physical interpretations of the type that I shall discuss here, we need to know about *censoring*.

Censoring

Consider an irreducible, finite-state discrete-time Markov chain whose states are partitioned into two sets E_1 and E_2. This induces a partitioning of its transition matrix T so that

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}.$$

The stationary distribution $\pi = (\pi_1, \pi_2)$ that satisfies $\pi T = \pi$ also satisfies

$$\pi_1 = \pi_1 \left[T_{11} + T_{12} \left(I - T_{22} \right)^{-1} T_{21} \right]$$

with

$$\pi_2 = \pi_1 T_{12} \left(I - T_{22} \right)^{-1}.$$

Censoring

Note that

$$(I - T_{22})^{-1} = \sum_{k=0}^{\infty} T_{22}^k$$

and so

$$\pi_1 \left[T_{11} + T_{12} (I - T_{22})^{-1} T_{21} \right] = \pi_1 \left[T_{11} + T_{12} \left[\sum_{k=0}^{\infty} T_{22}^k \right] T_{21} \right]$$

and we can interpret π_1 as the stationary distribution of the discrete-time Markov chain observed only when it is in E_1.

Similar comments can be made in the case where E_2 is infinite as long as $\sum_{k=0}^{\infty} T_{22}^k$ converges elementwise, which is the case when it leaves E_2 with probability one.

Censoring

In fact, we can say more:

It is not just the case that π_1 is the stationary distribution of the discrete-time Markov chain, censored so that it is observed only when it is in E_1, but

$$\left[T_{11} + T_{12} \left[\sum_{k=0}^{\infty} T_{22}^k \right] T_{21} \right]$$

is the transition matrix of this chain.

This is true even if the matrix is substochastic, in which case there is a positive probability that it may leave E_1 and not return.

Censoring

We can also observe that the (i, j)th entry of

$$[I - T_{22}]^{-1} T_{21} = \left[\sum_{k=0}^{\infty} T_{22}^{k}\right] T_{21}$$

is the probability that the Markov chain first enters E_1 in state j given that it started in state i of E_2.

Another Linear Algorithm

I claimed that it was hard to give a physical interpretation of Neuts' original algorithm.

We can, however, easily give a physical interpretation for a related algorithm due to Latouche.

Write the basic equation for G in a different way:

$$\begin{aligned}(I - A_1 - A_0 G)G &= A_2 \\ \Rightarrow G &= (I - A_1 - A_0 G)^{-1} A_2,\end{aligned}$$

and use the iteration

$$G_{n+1} = (I - A_1 - A_0 G_n)^{-1} A_2,$$

with $G_0 = 0$.

Another Linear Algorithm

The matrix $G_1 = (I - A_1)^{-1} A_2 = \left[\sum_{k=0}^{\infty} A_1\right] A_2$. Its (i,j)th entry is the probability that the chain hits level $k-1$ in finite time, does so in phase j, *and never reaches level* $k+1$, given that it starts in phase i of level k.

We can use induction to show that the (i,j)th entry of G_n is the probability that the chain hits level $k-1$ in finite time, and does so in phase j, *and never reaches level* $k+n$, given that it starts in phase i of level k, given that it starts in phase i of level k.

Thus, the successive iterates of this algorithm have the same physical interpretation as that of the matrix G, but with a linearly increasing taboo level.

A Quadratic Algorithm

In (1993), Latouche and Ramaswami proposed the *logarithmic-reduction algorithm*. This works by evaluating the expression

$$G = \sum_{\ell=0}^{\infty} \left[\prod_{i=0}^{\ell-1} U^i \right] D^\ell.$$

A Quadratic Algorithm

where the matrices U^ℓ and D^ℓ satisfy the recursion

$$U^{\ell+1} = \left[I - U^\ell D^\ell - D^\ell U^\ell \right]^{-1} \left[U^\ell \right]^2$$

and

$$D^{\ell+1} = \left[I - U^\ell D^\ell - D^\ell U^\ell \right]^{-1} \left[D^\ell \right]^2,$$

with $U^0 = (I - A_1)^{-1} A_0$ and $D^0 = (I - A_1)^{-1} A_2$.

A Quadratic Algorithm

If we let

$$\hat{G}_n = \sum_{\ell=0}^{n} \left[\prod_{i=0}^{\ell-1} U^i \right] D^\ell.$$

then, except when the QBD is null-recurrent, $\hat{G}_n$ converges to *G quadratically*. That is, there exists a constant $\gamma \in (0, 1)$ such that

$$\limsup_{n\to\infty} ||\hat{G}_n - G||^{1/2^n} = \gamma.$$

A Quadratic Algorithm

This algorithm also has a neat physical interpretation in terms of taboo probabilities.

The (i, j)th entry of the matrix $\hat{G}_n$ is the probability that the QBD will first enter level $k-1$ in phase j and does not visit any level higher than $k + 2^{n+1} - 2$ in between, given that it starts in phase i of level k.

Notice that the taboo level increases exponentially fast in terms of the number of iterates, which is consistent with the quadratic convergence of the algorithm.

A Quadratic Algorithm

The matrices $U^0 = (I - A_1)^{-1} A_0$ and $D^0 = (I - A_1)^{-1} A_2$ are, respectively the transition matrices of the discrete-time QBD derived from the original discrete-time QBD by observing it at the time points at which it changes level.

We can use induction to see that

$$U^{\ell+1} = \left[I - U^\ell D^\ell - D^\ell U^\ell \right]^{-1} \left[U^\ell \right]^2$$

and

$$D^{\ell+1} = \left[I - U^\ell D^\ell - D^\ell U^\ell \right]^{-1} \left[D^\ell \right]^2,$$

are the transition matrices of the discrete-time QBD derived from the original discrete-time QBD by observing it at the time points at which it hits levels of the form $k + m \times 2^{\ell+1}$.

A Quadratic Algorithm

So the summand on the right hand side of

$$G = \sum_{\ell=0}^{\infty} \left[\prod_{i=0}^{\ell-1} U^i \right] D^\ell.$$

takes into account sample paths that hit levels $k+1, k+3, \ldots, k+2^\ell-1$, all with level $k-1$ taboo, and then hits level $k-1$ with level $k+2^{\ell+1}-1$ taboo.

Further Work

The numerical analysis community has become interested in these problems.

In 1995, Bini and Meini adapted the *Cyclic Reduction Algorithm* to the calculation of G for processes of M/G/1 type. This algorithm uses a similar censoring idea to the logarithmic reduction algorithm, but organises the calculations slightly differently.

A number of speed-up features, such as transforming the matrices to move eigenvalues away from the unit circle and using Fast Fourier Transforms are now included in implementations. Benny Van Houdt maintains a web-site with state-of-the-art Matlab code.

Newton's Method

In deriving Latouche's linearly-convergent algorithm, we used the fact that

$$A_2 + A_1 G + A_0 G^2 = G$$

is equivalent to

$$G = (I - A_1 - A_0 G)^{-1} A_2.$$

One thing that we can do is apply Newton's method to the solution of this equation. We would expect this to lead to a quadratically-convergent algorithm.

Newton's Method

We obtain the sequence

$$G_{\mathrm{N}}^{(n+1)} - U^{(n)} A_0 G_{\mathrm{N}}^{(n+1)} U^{(n)} A_2 = U^{(n)} A_2 - U^{(n)} A_0 G_{\mathrm{N}}^{(n)} U^{(n)} A_2 \quad (\dagger)$$

where

$$U^{(n)} = (I - A_1 - A_0 G_{\mathrm{N}}^{(n)})^{-1}.$$

and

$$G_{\mathrm{N}}^{(0)} = 0.$$

The difficult part of implementing this is solving the Stein equation ($\dagger$) above for $G_{\mathrm{N}}^{(n+1)}$.

Newton's Method

In 1994, it was shown by Latouche that, for any initial matrix $G_{\mathrm{N}}^{(0)}$ with $0 \leq G_{\mathrm{N}}^{(0)} \leq G$, the sequence $G_{\mathrm{N}}^{(n)}$ converges monotonically and quadratically to G.

By transforming (†) into a standard linear system by concatenating the columns of $G_{\mathrm{N}}^{(n+1)}$ and writing the coefficient matrix as a direct sum involving $U^{(n)}A_0$ and $U^{(n)}A_2$, Latouche provided an algorithm for evaluating the sequence of matrices $\{G_{\mathrm{N}}^{(n)}\}$.

Using this transformation, he showed that each iteration of the algorithm has a complexity of order $O(m^6)$.

Newton's Method

Latouche tested Newton's algorithm against the linearly-convergent algorithm presented above and found that, while Newton's algorithm required up to an order of magnitude fewer iterations, it could take up to an order of magnitude longer in terms of computer time to calculate G to within a given tolerance.

After that time, it would be fair to say that the conventional wisdom in the matrix-analytic community was that the complexity of each iteration of Newton's method makes it uncompetitive with other algorithms. This attitude was only reinforced by the later discovery of the quadratically-convergent logarithmic-reduction algorithm.

Newton's Method

However, in 1992, Gardiner, Laub, Amato and Moler had provided a $O(m^3)$ algorithm for solving the Stein equation (†). This motivated us to revisit the question of how useful Newton's method is in this context.

We were also interested in the question of whether we can give a physical interpretation for Newton's method, in a similar vein to the physical interpretations discussed above for the linear and logarithmic reduction algorithms.

Physical Interpretation

As with the methods discussed above, the iterates $G_{\mathrm{N}}^{(n)}$ in Newton's Method contain the probabilities of certain sets of sample paths that start in level k and end in level $k-1$.

In the physical description that we gave above for the linear and quadratically-convergent algorithms, these sets were defined in terms of taboo levels.

In order to understand Newton's method, we need to look at the sample paths in a different way.

Physical Interpretation

Denote the set of sample paths taken into account in $G_{\mathrm{N}}^{(n)}$ by $\psi^{(n)}$.

The matrices $U^{(n)}A_0$ and $U^{(n)}A_2$ contain the probabilities of sets of sample paths that start in level k and end in levels $k+1$ and $k-1$ respectively. Denote these sets of sample paths by $\Phi_0^{(n)}$ and $\Phi_2^{(n)}$ *shifted to level* k.

Physical Interpretation

We have $U^{(0)} = (I - A_1)^{-1}$ and

$$\begin{aligned} G_{\mathrm{N}}^{(1)} &= U^{(0)} A_2 + U^{(0)} A_0 G_{\mathrm{N}}^{(1)} U^{(0)} A_2 \\ &= \sum_{\ell=1}^{\infty} \left((I - A_1)^{-1} A_0 \right)^{\ell-1} \left((I - A_1)^{-1} A_2 \right)^{\ell} \end{aligned}$$

where the second equation follows by repeatedly inserting the left hand side into the right hand side.

Physical Interpretation

So $\psi^{(1)}$ accounts for sample paths that start in level one, increase to some level ℓ, possibly remaining in any level along the way but never dropping back, and then decrease to level $k-1$, again possibly remaining in any level but never increasing.

Thus, the sample paths in $\psi^{(1)}$ are those that have a “single peak", no matter how high.

Sample Paths in $\psi^{(1)}$

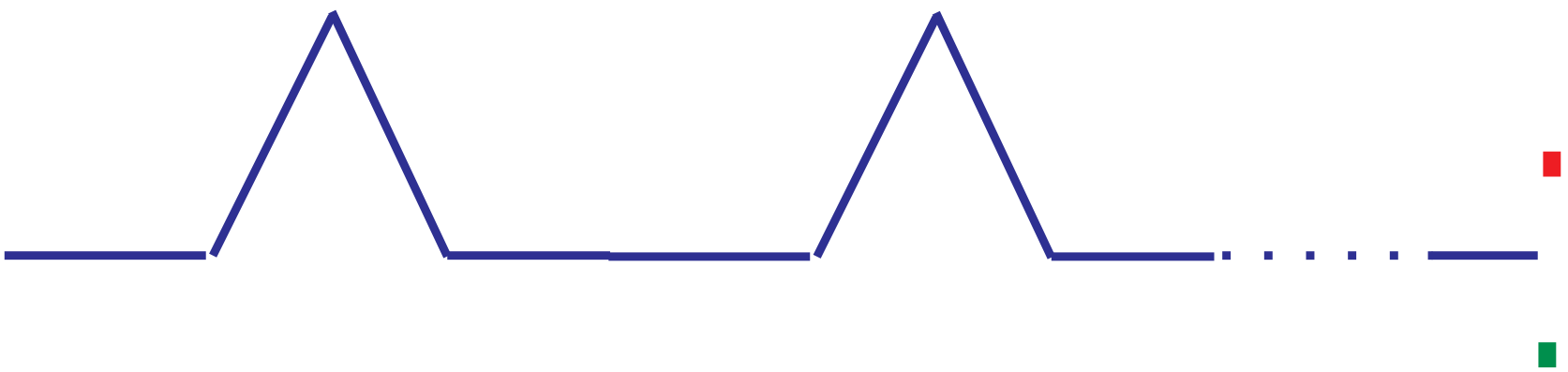
Physical Interpretation

The sets $\Phi_0^{(1)}$ and $\Phi_2^{(1)}$ contain sample paths taken into account by the matrices $U^{(1)}A_0$ and $U^{(1)}A_2$. We have

$$U^{(1)} = (I - A_1 - A_0 G_{\mathrm{N}}^{(1)})^{-1},$$

so $\Phi_0^{(1)}$ and $\Phi_2^{(1)}$ consist of sample paths that have any number of transitions between states at level k or “single peak" excursions from level k back to itself, followed respectively by a single transition to level $k+1$ and level $k-1$.

Sample Paths in $\Phi_0^{(1)}$ and $\Phi_2^{(1)}$



Physical Interpretation

The equation for $G_{\mathrm{N}}^{(2)}$ in terms of $G_{\mathrm{N}}^{(1)}$ and $U^{(1)}$ leads to

$$\begin{aligned} & G_{\mathrm{N}}^{(2)} - G_{\mathrm{N}}^{(1)} \\ = & \sum_{\ell=1}^{\infty} \left(U^{(1)} A_0\right)^{\ell-1} \left(U^{(1)} A_2\right)^{\ell} - \sum_{\ell=1}^{\infty} \left(U^{(1)} A_0\right)^{\ell} G_{\mathrm{N}}^{(1)} \left(U^{(1)} A_2\right)^{\ell}. \end{aligned}$$

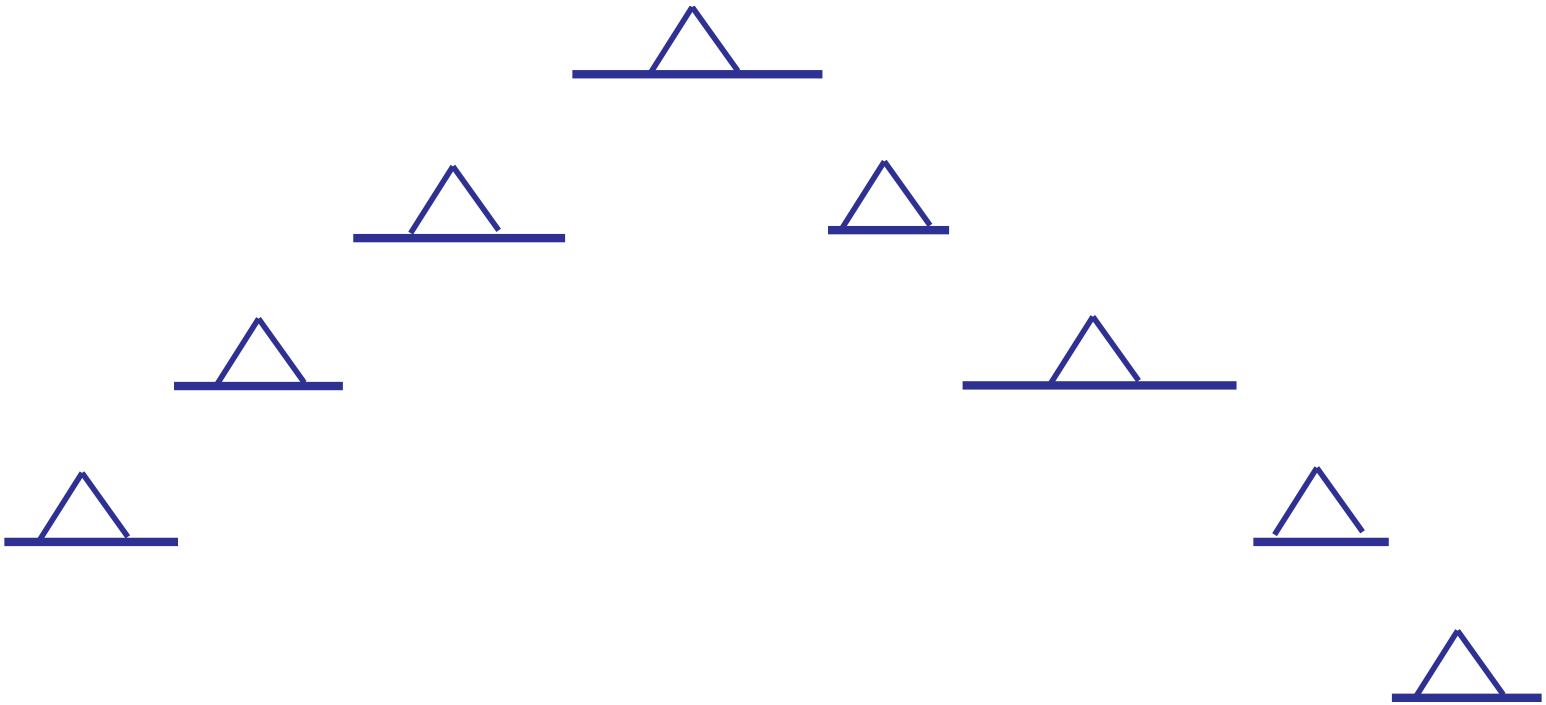
Physical Interpretation

So, sample paths in $\psi^{(2)}$ but not in $\psi^{(1)}$ are made up of a succession of sample paths that either stay at the same level or have “single peak" excursions upward, each shifted one level higher, up to some some level ℓ.

Then a succession of sample paths occurs that either stay at the same level or have “single peak" excursions upward, each shifted one level lower occurs, until the process drops to level $k-1$.

The subtraction of the second term on the right hand side of equation ensures that paths are not counted multiple times.

Sample Paths in $\psi^{(2)}$ but not in $\psi^{(1)}$



Physical Interpretation

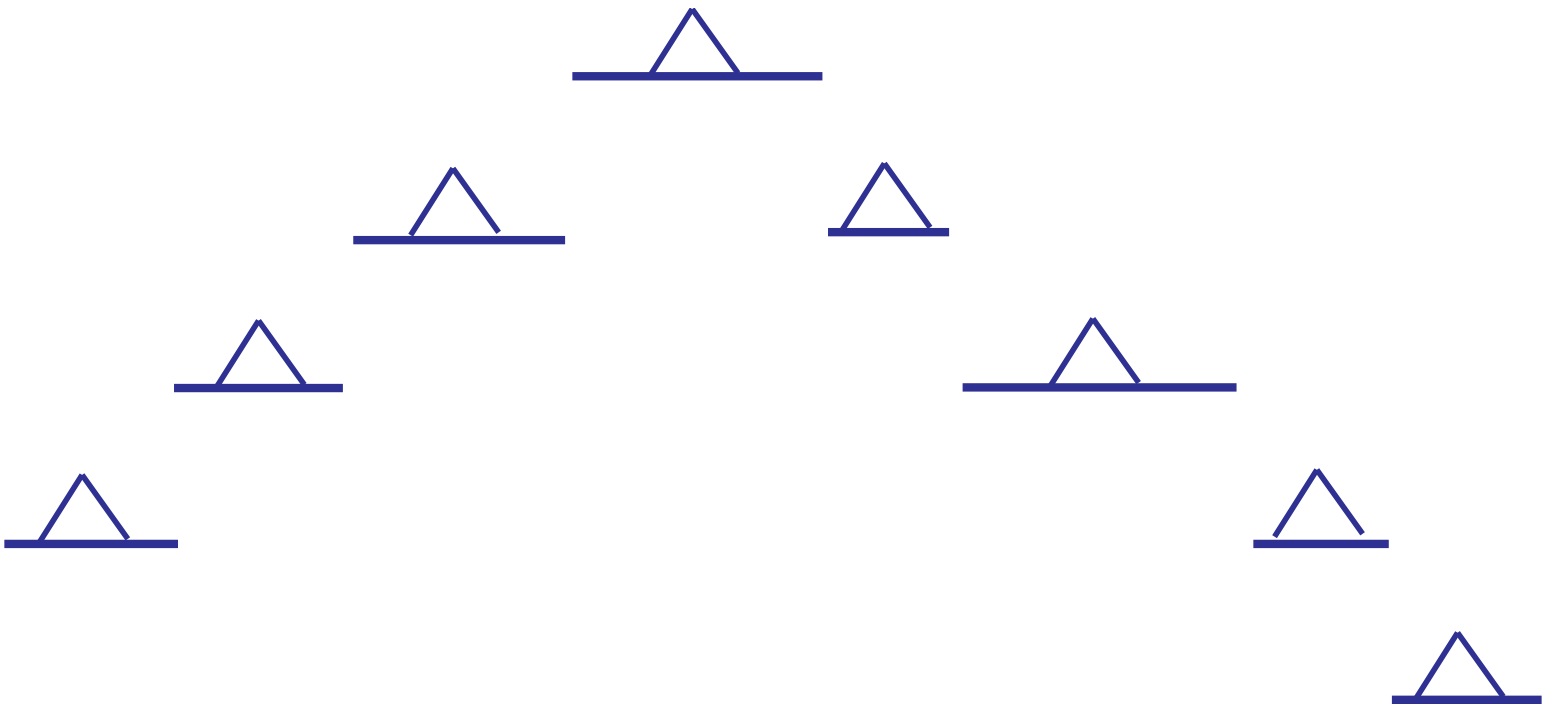
For general n,

$$
\begin{aligned}
& G_{\mathrm{N}}^{(n+1)} - G_{\mathrm{N}}^{(n)} \\
= & \sum_{\ell=1}^{\infty} \left(U^{(n)} A_0 \right)^{\ell-1} \left(U^{(n)} A_2 \right)^{\ell} - \sum_{\ell=1}^{\infty} \left(U^{(n)} A_0 \right)^{\ell} G_{\mathrm{N}}^{(n)} \left(U^{(n)} A_2 \right)^{\ell}.
\end{aligned}
$$

So, sample paths in $\Psi^{(n+1)}$ but not in $\Psi^{(n)}$ are made up of a succession of sample paths in $\Phi_0^{(n)}$, each shifted one level higher, up to some some level ℓ, whereupon a succession of sample paths in $\Phi_2^{(n)}$ occurs, each shifted one level lower, until the process drops to level $k-1$.

The subtraction of the second term on the right hand side of equation ensures that paths are not counted multiple times.

Sample Paths in $\psi^{(n+1)}$ but not in $\psi^{(n)}$



Physical Interpretation

So, instead of progressively including more sample paths by relaxing a taboo level, Newton's method progressively includes more and more complicated sample paths.

This happens in a 'fractal' way: basic units of paths at one iteration are the sets of paths that were accounted for in the previous iteration.

Very complicated paths are taken into account within a few iterations, which intuitively supports the fact that Newton's Method converges quadratically.