Final Report for grant FK 135711

titled

Mixing and information spread, Markov chains outside the comfort zone

The project being reported is in a special situation, as it has been decided to be cut short due to an upcoming, overlapping-interfering other project. As a result, the plans launched for 48 months have been closed after only 10 months. Certainly, research was initiated on multiple tracks, as detailed below, however, as it is well known, in mathematics both the expected duration of complete research output and their variance can be rather high. Consequently, in below we walk through the explicit results, alongside multiple tracks that are open-ended.

1 Completed research

Markov chains

One of the main tracks of research was targeting the understanding of mixing time improvement via non-reversible modifications of Markov chains. There are multiple problems that have been studied, which are still in process.

A reference example demonstrated before [3] used a cycle of n vertices prescribing deterministic steps in one orientation, then adding complete interconnections on a polynomially low number $n^c, 0 < c < 1$ randomly chosen vertices. A quadratic speedup was already observed in terms of the spectral gap compared to the symmetrized version with high probability. However, simulations showed already that using deterministic steps along the cycle is slightly suboptimal. Heuristically, adding a bit of laziness and/or backtracking will slow down the drift, but bring a bit of diffusion in exchange, which can balance favorably at first.

With Márton Beke, Márton Borbényi and Levente Szemerédi (all studying at Eötvös Loránd University) we performed a preliminary research to understand this perturbation analysis during a "Research Experience for Undergraduates" program, and succeeded to refine the previous understanding. Before, in [3] the only aim was establishing an upper bound on the absolute values of eigenvalues (except 1), in the process it was observed that they can be all found as roots of a certain random polynomial (except 0 with high multiplicity).

Now with examinations on lower bounds, it is now shown that $1 - |\lambda_2|$ is of the order of n^{c-1} (up to logarithmic corrections), moreover, for any $\epsilon \in (0,1)$, *all* non-zero eigenvalues will be larger than $1-\epsilon$ in absolute value asymptotically almost surely. Not worrying about exceptions but on average behavior, we find that for the n^c non-zero eigenvalues $n^{-c}\sum |\lambda_i|^r \to 1$ almost surely, for any positive integer r. However, suggesting an angular uniformity, at the same time we get $\sum \lambda_i^r \to 0$ almost surely. Indeed, it turns out that the non-zero part of the spectrum almost surely weakly converges to the uniform measure on the unit circle. Clearly these are powerful insights on the spectral structure, which will help understand and answer the motivating Markov chain perturbation problem afterwards.

A related model is a cycle of n vertices with a random matching on ϵn random vertices, which was the target studied together with David Levin (University of Oregon). The broad framework to build upon originates from the seminal work of [1], then later evolved to [7] which became a self-contained work but still highly technical. Our goal was to see what can be learned, what can be reused, what needs to be added.

There are non-trivial differences: in [7] the authors consider a random perfect matching to an arbitrary bounded degree graph, and work out mixing time and cut-off for the simple random walk on it. In our case, however, we do not have a perfect matching, we do not have symmetry, but the underlying graph is much simpler for the moment. Most importantly, we do not have an expander, but at least a local transience is true.

In particular, the model suggests that the typical and worst case mixing time will be different, that is, $t^x_{\text{mix}}(1/4) = (a + o(1)) \log n$ for asymptotically almost all starting points x, while $t^*_{\text{mix}}(1/4) = (b + o(1)) \log n$ for the worst starting point with $a < b$ and strict inequality, these values depending on ϵ and the local parameters of the chain. Cut-off is likely to carry through for typical starting points, for worst it is unclear at this moment.

At the current stage the research is based on having defined an associated tree structure that copes well with the drift which we name *layered tree*, which we can think of as (horizontal) copies of $\mathbb{Z}$, once in a while with a random extra edge hanging down, gaining a new copy of $\mathbb{Z}$ at the other end. Given a positive drift along each $\mathbb{Z}$ the overall asymptotic movement of a walker can be approximated nicely. Moreover, a joint generation of the random layered tree and the random matching can be performed which allows coupling of a walker on the tree and on the enhanced cycle as long as the latter seems locally tree-like. This exception, error probability is still being analyzed as it evolves with time. Note that the construction of the coupling is crucial, to explore and couple parts of the tree which are more likely in the trajectory of the walker. E.g., in the current construction being analyzed, for a certain vertex x on the explored portion of the tree, we look at the path from the root $\rho = v_0, v_1, \dots v_k = x$ and by denoting the horizontal displacements by $h(\cdot)$ if we find that $\max_i h(v_i) - h(x) > c \log \log n$ for some predetermined $c > 0$, we deny exploring further left from x. It would mean too many steps against the drift until reaching x then what is likely to occur in $O(\log n)$ steps, thus it is worth exploring elsewhere.

At a later step, the relaxation time of the completely revealed graph would be used for smoothing the distribution, but this is quite bad - of the order of $\log^2 n$, ruining the $\log n$ order - so a preparation is needed, cutting all long arcs to rather get $\log \log^2 n$, but then taking care of the mismatch.

As it can be seen, there are lot of small pieces, and several other pieces not mature enough to fill a complete puzzle yet.

There was also further work together with Andrea Ottolini (Stanford University) on understanding the mixing properties of a Gibbs sampler Markov chain in $[0,1]^d$ which besides its own interest has an external motivation from Bayesian statistics. Almost exchangeability can be expressed by a density proportional to $\exp(-A^2 \sum (x_i - x_j)^2)$, and even generalized to $\exp(-A^2 \sum c_{ij}(x_i - x_j)^2)$. Starting from our result [6] showing mixing time of order A^2 (with constants depending on c_{ij}), we embarked on the follow-up question Bayesian statisticians would tell, when data is available, and a posterior distribution is to be treated. This results in a distribution with density proportional to

$$\exp(-A^2 \eta \sum c_{ij}(x_i - x_j)^2 - A^2(1 - \eta) \sum d_i(x_i - x_i^*)^2)$$

for some reference point $x^* \in [0,1]^d$. Our studies already revealed that we will have to combine random matrix products, quantify precisely the distances of various Gaussians and deal with the conditioning of the hypercube. The hypercube conditioning was clarified at first, other parts were not entirely completed in the time time available, but some derandomized simplifications suggested a much faster mixing than the original diffusive A^2 of [6].

An additional topic of focus was random walks in random environments. In such problems, a major goal is to relax requirements as possible, mostly in terms of dependence and in terms of finiteness. Our recent work with Miklós Rásonyi (Rényi Institute) [4] was dealing with scalar valued fractional volatility models for financial assets, a stochastic modeling concept in modern financial mathematics which takes into account the non-Markovian nature of volatilities, where the price process develops as a time-inhomogeneous Markov process given the volatility process. Originally it was shown that there exist a stationary measure for the price and convergence takes place in probability.

We now upgraded our analysis to include multidimensional assets, which is a clear call from applications, where the interwoven development of various products create a non-trivial structure. Both cases build upon a finely crafted coupling construction of a stationary and a yet-to-be stationary process through a family of random transformations (which have to carefully parametrized due to the random environment).

Distributed averaging

The other main direction was related in spirit, but with an application inspiration, on the analysis of the distributed averaging algorithm schemes named push-sum (or ratio consensus, weighted gossip). A major body of work jointly with László Gerencsér (SZTAKI) identifying precisely the convergence rate of such algorithms in a quite general case has been iterated and refined and is now accepted and available online [2] in an excellent publication outlet. The strength of the work is that it only requires stationary ergodic updates rather than any stronger independence notion and the acquired exact bounds, which are expressed through the Lyapunov exponents of the update matrix series.

A follow-up work has been carried out for simplification. The motivation was that despite the precise expressions above, Lyapunov exponents are hard to compute in general, and are asymptotic quantities by definition, so it could be helpful to get a simpler bound, even if less precise. This is carried out for i.i.d. series [5], presenting a convergence rate bound that only uses the expectation of a transformation of a single random update matrix, avoiding an asymptotic viewpoint. We also confirm that the resulting bound is always non-trivial in the sense that a strictly negative exponential rate is stated. Some numerical tests affirmed that the bound can nicely get close to the true rate. Note that a general control on the (relative) error cannot be established by the known unapproximability results of the Lyapunov exponents.

Work did not stop here, together with Miklós Kornyik (Rényi Institute) we wanted to get more hands-on bounds based on the computable expression above. The goal was to focus on natural update schemes and relate them to fundamental, accessible quantities. To be more specific on ratio consensus and our goal, along a vector of inputs x_0 a "weight" vector $w_0 = 1$ is initialized, then the updates at every time step can be described by a non-negative column-stochastic matrix A_t used to map both the primary value and the auxiliary weight variables:

$$x_t = A_t x_{t-1}, \qquad w_t = A_t w_{t-1},$$

with the estimate of the average being x_t/w_t at time t. The process to be understood was chosen when agents communicate along a fixed connected graph, and at every step each agent (thinking of acting for the coordinates) sends a fixed portion q to a randomly chosen neighbor. Our ideal was to find a bound in terms of the spectral properties of the graph.

We have found satisfying bounds for the case of transitive graphs, showing that

$$\limsup_t \frac{1}{n} \log \left\| \frac{x_t}{w_t} - \bar{x} \right\| \leq \frac{1}{2} \log[(1-q+q\lambda_2)^2 + q^2(1-\lambda_2^2)] \approx -(1-\lambda_2)(q-q^2),$$

where $\bar{x}$ is the average of the initial values x_0 and λ_2 is the second largest eigenvalue of the symmetric random walk on the graph. Here the first term corresponds to a natural linear improvement of the rate as q is increased (near 0) while the second one is a corrective factor, notice however that the $(1-\lambda_2^2)$ coefficient drives it rather small for graphs that are non-expanding. Using the $\log(1-x) \approx -x$ approximation the expression becomes even simpler near $q = 0$, Simulations confirm that the bound follows well the numerical rates not only for negligible q-s, but easily up to 0.4-0.6 depending on the graph tested.

Our goal was to extend the bound above to more general classes of graphs, which seems surprisingly harder, even for arbitrary regular graphs analogous bounds could not yet be carried over.

Additional scientific activity

Besides research itself, there was activity on building and maintaining connections, and exchanging our thoughts. The pandemic situation did decrease the opportunities, still it was possible to travel to Louvain-la-Neuve, Belgium, to get a new boost for the long-term collaborations there, to share recent research results - including discussions on [2], [5] - to get more involved. Participation at the European Congress of Mathematics could have been even more fruitful with a fully on-site conference, still helped to build new connections to the Polish Mathematical Society, hopefully with a more active future.

At Rényi Institute, I served as the scientific secretary, helping the management and the director on semi-practical issues, e.g., making sure seminar announcements are reaching all possibly interested.

Perspectives

Most perspectives are rather natural for the open-ended tracks. The PI needs a discussion with the numerous peers listed above, to evaluate the status of the ongoing tracks. Assuming a positive and optimistic outcome and that both parties are willing to spend time again on it, they can restart investigation, building on lemmas and partial results previously obtained. To summarize and highlight for convenience, along the two main themes we have questions to answer:

Markov chains:

- Modify the drift to improve mixing on a cycle of n vertices with interconnections on a polynomially low number $n^c, 0 < c < 1$ randomly chosen vertices.
- Determine mixing and cutoff on a cycle of n vertices with a random matching on ϵn random vertices.
- Determine mixing and cutoff for the Gibbs sampler on $[0, 1]^d$ for the posterior distribution for almost exchangeability, i.e., when the target distribution gets more away from the homogeneous near-diagonal reference term.

Push-sum:

- Establish bounds on the convergence rate of push-sum schemes via a spectral description of the connectivity graph, for arbitrary connected regular graphs, then for more general graph classes.

Moreover, given the new project the PI is leading, the knowledge and willingness to cooperate gathered will also be channeled there. Research topics will be shifted to be infused more with optimization themes, with a perspective of not only analyzing specific dynamics, but towards exploring the neighborhood for the best performance.

References

[1] N. Berestycki, E. Lubetzky, Y. Peres, and A. Sly, *Random walks on the random graph*, The Annals of Probability, 46 (2018), pp. 456–490.

[2] B. Gerencsér and L. Gerencsér, *Tight bounds on the convergence rate of generalized ratio consensus algorithms*, IEEE Transactions on Automatic Control, (2021). Available online.

[3] B. Gerencsér and J. M. Hendrickx, *Improved mixing rates of directed cycles by added connection*, Journal of Theoretical Probability, 32 (2019), pp. 684–701.

[4] B. Gerencsér and M. Rásonyi, *Invariant measures for fractional stochastic volatility models.* arXiv:2002.04832, 2020.

[5] B. Gerencsér, *Computable convergence rate bound for ratio consensus algorithms.* arXiv:2104.04802, 2021.

[6] B. Gerencsér and A. Ottolini, *Rates of convergence for gibbs sampling in the analysis of almost exchangeable data.* arXiv:2010.15539, 2020.

[7] J. Hermon, A. Sly, and P. Sousi, *Universality of cutoff for graphs with an added random matching*, arXiv:2008.08564, (2020).