10 Sparse matrices

Up to this point we have only analyzed random matrices whose entries are all (with high probability) nonzero random numbers. However, matrices with many zero components are often interesting. They usually arise because we consider some system whose interactions are defined by a network or lattice, where the adjacency matrix of the lattice C defines its connectivity: $C_{ij} = 1$ if i and j interact and $C_{ij} = 0$ otherwise. Then, a matrix of *random* interactions would have the form

$$H_{ij} = C_{ij}J_{ij} + H_{ii}\delta_{ij} \tag{1}$$

for J a, e.g., GOE matrix and allowing for autonomous diagonal entries, or 'on-site' interactions. In general such systems cannot be solved directly, but when the network is tree-like (few loops whose size grows with N) there are techniques to write down arbitrary-precision numeric solutions.

We start again from the block decomposition (or Schur complement formula) of the resolvent matrix

$$G_{ij}(z) = (zI - H)^{-1}_{ij} \tag{2}$$

where as always the final resolvent is given by

$$G(z) = \frac{1}{N}\sum_{i=1}^{N}\overline{G_{ii}(z)} \tag{3}$$

The decomposition where the ith row and column is isolated from the rest of the matrix is given in general by

$$\begin{aligned} G_{ii}(z)^{-1} &= G_{ii}^{-1}(z) - \sum_{j,k\neq i} G_{ij}^{-1}(z)G_{jk}^{(i)}(z)G_{ki}^{-1}(z) \\ &= z - H_{ii} - \sum_{j,k\neq i} H_{ij}G_{jk}^{(i)}H_{ki} \end{aligned} \tag{4}$$

where $G^{(i)}$ is the resolvent for the matrix H with the ith row and column removed. When studying fully connected systems, we quickly argued

1. $\overline{G_{ii}(z)^{-1}} = (\overline{G_{ii}(z)})^{-1}$ using the self-averaging of the resolvent, and
2. $\overline{G_{ii}(z)} = \frac{1}{N}\sum_{j\neq i}\overline{G_{jj}^{(i)}} = G(z)$, arguing that the resolvent of only the ith site and the resolvent of the matrix without row i are both on average equal to the resolvent of the whole matrix.

Both of these arguments fail in the sparse case. Before, each row 'interacted' with $N-1$ others, and we could rely on these properties that arise from the law of large numbers. However, now each row 'interacts' with only a few others, and such conveniences no longer apply.

We can see the effect of this by letting the network nature of the matrix entries be explicit in our formula. We use the notation

$$\partial i = \{j \mid C_{ij} = 1\} \tag{5}$$

to describe the set of neighboring sites to i. Then we can write

$$\begin{aligned} G_{ii}(z)^{-1} &= z - H_{ii} - \sum_{j,k\neq i} C_{ij}C_{ki}J_{ij}G^{(i)}_{jk}(z)J_{ki} \\ &= z - H_{ii} - \sum_{j,k\in\partial i} J_{ij}G^{(i)}_{jk}(z)J_{ki} \end{aligned} \tag{6}$$

Now we make the crucial assumption unique to the cavity method on sparse networks: that with the site i removed, all properties attached to sites $j, k \in \partial i$ are independent from each other. This is an exact assumptions on a tree, since removing the site i literally disconnects its neighbors in the network, and the matrix of interactions $C^{(i)}$ can be blockwise decomposed into blocks associated with each former neighbor. In particular, this means that eigenvectors of $H^{(i)}$ and therefore $G^{(i)}$ only have support on one or the other site: if $v^{(i)}_{lj} \neq 0$ for the $j \in \partial i$ component of some eigenvector $\boldsymbol{v}^{(i)}_l$ of $H^{(i)}$, then $v^{(i)}_{lk} = 0$ for the other neighbors $k \in \partial i$. We therefore have for $j, k \in \partial i$ that

$$G^{(i)}_{jk}(z) = (zI - H^{(i)})^{-1}_{jk} = \sum_{l=1}^{N} \frac{v^{(i)}_{lj}v^{(i)}_{lk}}{z - x^{(i)}_l} = \sum_{l=1}^{N} \frac{(v^{(i)}_{lj})^2\delta_{jk}}{z - x^{(i)}_l} = \delta_{jk}G^{(i)}_{jj}(z) \tag{7}$$

that is, the resolvent matrix for the lattice with i removed is diagonal among the neighbors of site i. We then have

$$G_{ii}(z)^{-1} = z - H_{ii} - \sum_{j\in\partial i} J^2_{ij}G^{(i)}_{jj}(z) \tag{8}$$

The objects $G^{(i)}_{jj}$ are called the *cavity Green functions* since they are the diagonal resolvent entries (Green functions) for the network with a cavity or hole in it due to the absence of i. The diagonal entries $G_{ii}(z)$ are sometimes called the *single-site marginal Green functions*.

The payoff of this construction is that we can write a self-consistent set of equations for the cavity functions. We do this by repeating the above argument point for point, but now for the matrix $G^{(i)}$: this time, use the block inversion formula isolating the jth row of $G^{(i)}$ for $j \in \partial i$. Then

$$G^{(i)}_{jj}(z)^{-1} = z - H_{jj} - \sum_{k\in\partial j\backslash i} J^2_{jk}G^{(i,j)}_{kk} \tag{9}$$

where the notation $\partial j\backslash i$ means that we consider the set ∂j with i removed (since the matrix we are considering has i removed), and $G^{(i,j)}$ is the resolvent for the matrix with rows and columns i and j removed. Now, the crucial insight is that given the tree-like assumption, the cavity Green function attached to site k doesn't know that the site i was

removed from the network, since the site i is only connected to k via j, which is also removed. Therefore, we can write $G^{(i,j)}_{kk} = G^{(j)}_{kk}$, and finally

$$G^{(i)}_{jj}(z)^{-1} = z - H_{jj} - \sum_{k \in \partial j \setminus i} J^2_{jk} G^{(j)}_{kk} \tag{10}$$

which is a closed set of equations for the cavity Green functions.

The set of equations for the single-site marginal and cavity Green functions is not so easy to solve in general, notably because they form a system of coupled equations for N different values of the marginal functions and $2E$ for number of network edges E different values of the cavity functions. For a specific network with a specific set of couplings and diagonal elements, the equations can be solved by iteration for a given value of z:

1. Start with a random guess for $G^{(i)}_{jj,0}(z)$ for every bond (i, j).

2. Use the equation above to generate new guesses $G^{(i)}_{jj,n+1}(z)$ from the previous guess by

$$G^{(i)}_{jj,n+1}(z)^{-1} = z - H_{jj} - \sum_{k \in \partial j \setminus i} J^2_{jk} G^{(j)}_{kk,n}(z) \tag{11}$$

3. Repeat the process of new guesses until the cavity Green functions have converged.

4. Calculate the single-site marginals from (8).

This algorithm appears generally in inference and statistics problems on networks and is called *message passing*. In practice, one in interested in the spectral density, and so chooses $z = x - i\epsilon$ for some small but nonzero ϵ and a solves the problem on a grid of x. Since for nearby x the problem should have nearby solutions, one can initialize the algorithm with the output of a previous step.

It might seem silly to approach the spectral density this way, since for a matrix with specific interactions one can just exactly diagonalize it. However, this algorithm is much faster for large matrices, since typically eigenvalue solving takes N^3 runtime. Each iteration of this algorithm takes order $E \leq 2pN$ for networks with maximum vertex degree p, and convergence is typically only in order-one steps. Therefore, one often has *linear-time* convergence for message passing, compared to cubic for exact diagonalization.

The equations can be solved exactly when the network is homogenous and deterministic. Suppose that in our network every vertex has degree p, $J_{ij} = \frac{1}{\sqrt{2}} J$ for all i, j, and $H_{jj} = 0$. Such a network is called an *random regular graph*. Assuming they all take the same value $G^{(i)}_{jj}(z) = G_c(z)$, the equation for the cavity Green functions becomes

$$G_c(z)^{-1} = z - \frac{1}{2}(p-1) J^2 G_c(z) \tag{12}$$

where we have used the fact that site j has $p-1$ neighbors excluding i. This can be solved exactly to yield

$$G_c(z) = \frac{1}{(p-1)J^2}\left(z \pm \sqrt{z^2 - 2(p-1)J^2}\right) \tag{13}$$

Then the equation for the single-site marginal Green functions, assuming they are all identical again, yields

$$\begin{aligned} G(z)^{-1} &= z - \frac{1}{2}pJ^2 G_c(z) \\ &= z - \frac{p}{2(p-1)}\left(z \pm \sqrt{z^2 - 2(p-1)J^2}\right) \end{aligned} \tag{14}$$

where now we use the fact that site i has p neighbors. The spectral density implied by this resolvent can be compared with that resulting from generating random regular graphs of degree p, and they match well for large graphs, even though random regular graphs are not trees![1] Random regular graphs have an important property that the expected size of loops grows with N sufficiently large that the tree-like assumption of independence of the cavity Green functions on different neighbors holds.

An interesting limit is that of high-connectivity, where p approaches infinity with J scaled to keep things finite. The good scaling for J is $J = p^{-\frac{1}{2}}$, which gives as $p \to \infty$

$$G(z)^{-1} = \frac{1}{2}\left(z \mp \sqrt{z^2 - 2}\right) \tag{15}$$

which gives the semicircle distribution! In fact this distribution can be recovered from the cavity equations even for random J in this limit, and it should make sense: with high connectivity, the law-of-large-numbers assumptions that went into the fully connected cavity calculation become correct again.

However, in this course and in life we are often interested in the spectral density *averaged over the random ensemble*. How can we use these equations to arrive at this? When J and the diagonal of H and even the connectivity of the network C are random, we must understand the cavity equations as *distributional* equations, in the specific sense that they tell us the probability distribution of $G^{(i)}_{jj}$ must obey

$$\begin{aligned} p(G^{(i)}_{jj}, d) = \sum_C P(C) \int \prod_{k \in \partial_C j \setminus i} \left(dp(J_{jk})\, dp(G^{(j)}_{kk}, |\partial_C k|)\right) \int dp(H_{jj}) \\ \times \delta\left((G^{(i)}_{jj})^{-1} - z + H_{jj} + \sum_{k \in \partial_C j \setminus i} J^2_{jk} G^{(j)}_{kk}\right) \end{aligned} \tag{16}$$

where $dp(x) = dx\, p(x)$ is a compact way of writing the measure over the probability distribution p for x. This is a functional equation for the joint probability that the cavity

[1] In fact this works better for random regular graphs than finite trees, because finite trees have a boundary of degree-one leaves whose volume dwarfs that of the rest of the tree.

function and its degree have specific values, and if we had that probability we could calculate the probability of the single-site marginals by solving

$$p(G_{ii}, d) = \sum_C P(C) \int \prod_{k \in \partial_C i} \left(dp(J_{ij})\, dp(G^{(i)}_{jj}, |\partial_C j|) \right) \int dp(H_{ii}) \qquad (17)$$
$$\times \delta \left((G_{ii})^{-1} - z + H_{ii} + \sum_{j \in \partial_C i} J_{ij}^2 G^{(i)}_{jj} \right)$$

These really look like hopeless equations, but there is a simple and fast algorithm for solving them numerically called *population annealing*, which is also quite general for self-consistent probability problems like this. The algorithm works by creating a population of M candidate samples from the distribution, and then repeatedly replace them with new samples generated by the consistency relation. If the distribution of samples (not the individual samples!) converges, then the result is distributed with the desired probability distribution. The operation is:

1. Generate a population of M cavity Green functions G_i at random.
2. Draw a new degree d from the distribution of degrees in the network, and h from the distribution of diagonal elements.
3. Draw d samples from the population, with $\sigma \subset \{1, \ldots, M\}$, $|\sigma| = d$, and for each of them generate J_i from the distribution of off-diagonal weights.
4. Generate a new cavity Green function G with
$$G^{-1} = z - h - \sum_{i \in \sigma} J_i^2 G_i \qquad (18)$$
5. Replace one of the existing elements of the population with the new one at random
6. Repeat 2–5 until the distribution of G has converged.

Then, the distribution of single-site marginals can be computed by repeatedly sampling the converged population in a similar way, i.e., drawing d from the degree distribution and calculating the marginal picking d neighbors at random from the population. Finally, the expected value of the resolvent is given by taking the mean of the sample of single-site marginal Green functions.

This algorithm effectively works at infinite size, and its accuracy scales with the size of the population M. The main source of systematic error in population annealing is that features of the probability distribution can only be resolved if they correspond to a PDF greater than M^{-1}, so tail behavior can be difficult to resolve. Unfortunately, tail behavior is an important part of phenomena like localization, so work on resolving it is of concern. However, there are sophisticated extensions of population annealing that better resolve small-probability (or large-deviation) parts of the distribution of cavity fields, and so are suitable for treating such problems.