The Perron–Frobenius theorem and the Hilbert metric

Vaughn Climenhaga

April 17, 2013

In the last post, we introduced basic properties of convex cones and the Hilbert metric. In this post, we look at how these tools can be used to obtain an explicit estimate on the rate of convergence in the Perron–Frobenius theorem.

1 Perron–Frobenius theorem

We start by stating a version of the Perron–Frobenius theorem. Let A be a $d \times d$ stochastic matrix, where here we use this to mean that the entries of A are non-negative, and every column sums to 1: $A_{ij} \in [0,1]$ for all i,j, and $\sum_{i=1}^d A_{ij} = 1$ for all j. Thus the columns of A are probability vectors.

Such a matrix A describes a weighted random walk on d sites: if the walker is presently at site j, then A_{ij} gives the probability that he will move to site i at the next step. Thus if we interpret a probability vector v as giving the probability of the walker being at site j with probability v_j, then $v \mapsto Av$ gives the evolution of this probability under one step of the random walk.

Now one version of the Perron–Frobenius theorem is as follows: If A is a stochastic matrix with $A > 0$ (that is, $A_{ij} > 0$ for all i,j), then there is exactly one probability vector π that is an eigenvector for A. Moreover, the eigenvalue associated to this eigenvector is 1, the eigenvalue 1 is simple, and all other eigenvalues have modulus < 1. In particular, given any $v \in [0,\infty)^2$ we have $A^n v \to \pi$ exponentially quickly.

The eigenvector π is the stationary distribution for the random walk (Markov chain) given by A, and the convergence result states that any initial

distribution converges to the stationary distribution under iteration of the process.

The assumption that $A > 0$ is quite strong: for the random walk, this says that the walker can get from any site to any other site in a single step. A more general condition is that A is *primitive*: that is, there exists $N \in \mathbb{N}$ such that $A^N > 0$. This says that there is a time N such that by taking N steps, the walker can get from any site to any other site. The same result as above holds in this case too.

In fact, the result holds in the even more general case when A is *irreducible*: for every i, j there exists N such that $(A^N)ij > 0$. This says that the walker can get from every site to every other site, but removes the assumption that there is a single time N that works for all site. For example, consider a random walk on a chessboard, where the walker is allowed to move one square horizontally or vertically at each step. Then for a sufficiently large even value of N, the walker can get from any white square to any other white square, but to get to a black square requires an odd value of N.

2 Rate of decay using convex cones

As stated above, the Perron–Frobenius theorem does not give any result on the rate with which $A^n v$ converges to π. One way to give an estimate on this rate is to use convex cones and the Hilbert metric, which were discussed in the last post.

2.1 A cone and a metric

Let $\mathcal{C}$ be the convex cone $[0, \infty)^d \subset \mathbb{R}^d$. We want an estimate on the diameter of $A(\mathcal{C})$ in the Hilbert metric $d_\mathcal{C}$. Recall that this metric is given by $d_\mathcal{C}(v, w) = \log(\beta/\alpha)$, where

$$\beta = \inf\{\mu > 0 \mid \mu v - w \in \mathcal{C}\},$$
$$\alpha = \sup\{\lambda > 0 \mid w - \lambda v \in \mathcal{C}\}.$$

Another way of interpreting the cone $\mathcal{C}$ is in terms of the partial order it places on V, which is given by $v \preceq w \Leftrightarrow w - v \in \mathcal{C} \cup \{0\}$. We see that β and α can be characterised as

$$\alpha = \sup\{\lambda \mid \lambda w \preceq v\}, \qquad \beta = \inf\{\mu \mid v \preceq \mu w\}.$$

In our present example, we see that the cone $\mathcal{C} = [0, \infty)^d$ induces the partial order $v \preceq w \Leftrightarrow v_i \leq w_i \ \forall i$. Thus

$$\alpha = \sup\{\lambda \mid \lambda w_i \leq v_i \ \forall i\} = \min_{1 \leq i \leq d} \frac{v_i}{w_i}, \tag{1}$$

and similarly $\beta = \max_{1 \leq i \leq d} \frac{v_i}{w_i}$.

2.2 Diameter of $A(\mathcal{C})$

Now we need to determine the diameter Δ of $A(\mathcal{C})$ in the Hilbert metric $d_{\mathcal{C}}$. If $\Delta < \infty$, then the theorem of Birkhoff from the previous post will imply that $d_{\mathcal{C}}$ contracts distances by a factor of $\tanh(\Delta/4) < 1$.

Let e_i be the standard basis vectors in $\mathbb{R}^d$. Because $d_{\mathcal{C}}$ is projective we can compute Δ by considering $d_{\mathcal{C}}(Av, Aw)$ where $\sum v_i = \sum w_j = 1$. Using the triangle inequality, we have

$$d_{\mathcal{C}}(Av, Aw) = d_{\mathcal{C}}\left(A \sum v_i e_i, A \sum w_j e_j\right) = d_{\mathcal{C}}\left(\sum v_i (A e_i), \sum w_j (A e_j)\right)$$
$$\leq \sum_{i,j} v_i w_j d_{\mathcal{C}}(A e_i, A e_j) \leq \max_{i,j} d_{\mathcal{C}}(A e_i, A e_j),$$

so it suffices to consider $d_{\mathcal{C}}(A e_i, A e_j)$ for $1 \leq i, j \leq d$. But $A e_i$ is just the ith column of the matrix A, so writing $A = [v^1 \cdots v^n]$, where v^i is the ith column vector, we see that

$$\Delta \leq \max_{i,j} d_{\mathcal{C}}(v^i, v^j). \tag{2}$$

2.3 Contraction under multiplication by A

Now we have a very concrete procedure for estimating the amount of contraction in the $d_{\mathcal{C}}$ metric under multiplication by A:

1. estimate Δ using (2) and the expression for $d_{\mathcal{C}}$ in (1) and the discussion preceding it;

2. get a contraction rate of $\tanh(\Delta/4) < 1$.

From (1) and the discussion preceding it, the distance $d_{\mathcal{C}}(v^i, v^j)$ is given as

$$d_{\mathcal{C}}(v^i, v^j) = \log \beta - \log \alpha = \log \left(\max_{1 \leq k \leq d} \frac{v_k^i}{v_k^j} \cdot \max_{1 \leq k \leq d} \frac{v_k^j}{v_k^i} \right). \tag{3}$$

3

Let $\Lambda = \tanh(\Delta/4)$. To write an explicit estimate for Λ, we use

$$\Lambda = \frac{e^{\Delta/4} - e^{-\Delta/4}}{e^{\Delta/4} + e^{-\Delta/4}} = \frac{1 - e^{-\Delta/2}}{1 + e^{-\Delta/2}} \leq \frac{1 - s}{1 + s}, \tag{4}$$

where $s < 1$ is any estimate we can obtain satisfying $e^{-\Delta/2} \geq s$. From (3) and (2), we have

$$e^{-\Delta/2} \geq \max_{i,j} \sqrt{\min_k \left(\frac{v_k^i}{v_k^j}\right) \min_k \left(\frac{v_k^j}{v_k^i}\right)} =: s. \tag{5}$$

This allows us to obtain estimates on $d_{\mathcal{C}}(A^n v, A^n w)$. However, we want to estimate $d(A^n v, A^n w)$ in a more familiar metric, such as one coming from a norm. We can relate the two by observing that if $v, w \in (0, 1]^d$, then

$$d_{\mathcal{C}}(v, w) = \log \max_k \left(\frac{v_k}{w_k}\right) + \log \max_k \left(\frac{w_k}{v_k}\right)$$
$$\geq \max_k |\log v_k - \log w_k| \geq \max_k |v_k - w_k| = \|v - w\|_{L^\infty},$$

where the last inequality uses the fact that log has derivative ≥ 1 on $(0, 1]$. Since A maps the unit simplex to itself (because A is stochastic), we see that

$$\|A^n v - A^n w\|_{L^\infty} \leq d_{\mathcal{C}}(A^n v, A^n w) \leq C\Lambda^n, \tag{6}$$

where Λ is given by (4) and (5), and where we can take either $C = d_{\mathcal{C}}(v, w)$ or $C = \Delta/\Lambda$ (since $d_{\mathcal{C}}(Av, Aw) \leq \Delta$), whichever gives the better bound. Since all norms on $\mathbb{R}^d$ are equivalent, we have a similar bound in any norm.

3 Nonnegative matrices

The analysis in the previous section required A to be positive ($A_{ij} > 0$ for all i, j). A more general condition is that A is nonnegative and primitive: that is, $A_{ij} \geq 0$ for all i, j, and moreover there exists N such that $A^N > 0$.

If $A_{ij} = 0$ for some i, j, then it is easy to see from the calculations in the previous section that $A(\mathcal{C})$ has infinite diameter in the Hilbert metric, so the above arguments do not apply directly. However, they do apply to A^N when $A^N > 0$, and so we fix N for which this is true, and we obtain $\Lambda < 1$ such that $d_{\mathcal{C}}(A^N v, A^N w) \leq \Lambda d_{\mathcal{C}}(v, w)$ for all $v, w \in \mathcal{C}$.

Moreover, let $L \in \mathbb{R}$ be such that $\|A^r\| \leq L$ for all $0 \leq r < N$. Then for any $n \in \mathbb{N}$ we can write $A^n = A^{kN+r}$ for some $0 \leq r < N$, so that

$$\|A^n v - A^n w\| = \|A^r(A^{kN}v - A^{kN}w)\| \leq LC\Lambda^k,$$

where C is as in (6). Thus we conclude that asymptotically, $A^n v$ approaches the eigenvector with contraction rate $\Lambda^{1/N}$.

To see this in action, consider a Markov chain with transition matrix

$$A = \begin{pmatrix} \frac{1}{2} & 1 \\ \frac{1}{2} & 0 \end{pmatrix}.$$

That is, from the first state the walker transitions to either state with probability $1/2$, while from the second state the walker always returns to the first state. Since the transition from the second state to itself is forbidden, $A(\mathcal{C})$ has infinite diameter. However, the two-step transition matrix is

$$A^2 = \begin{pmatrix} \frac{3}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} \end{pmatrix},$$

for which we can compute

$$s = \sqrt{\frac{1/4}{1/2} \cdot \frac{1/2}{3/4}} = \frac{1}{\sqrt{6}} \quad \Rightarrow \quad \Lambda \leq \frac{\sqrt{6}-1}{\sqrt{6}+1}.$$

Thus the estimate on A^2 gives us a definite rate of contraction, which the estimate from A does not.

It can be useful to use the estimate on A^N even when $A > 0$. For example, if we consider the Markov chain with transition matrix

$$A = \begin{pmatrix} \frac{1}{5} & \frac{9}{10} \\ \frac{4}{5} & \frac{1}{10} \end{pmatrix},$$

then we have

$$s = \sqrt{\frac{1/5}{9/10} \cdot \frac{1/10}{4/5}} = \sqrt{\frac{2}{9} \cdot \frac{1}{8}} = \frac{1}{6} \quad \Rightarrow \quad \Lambda \leq \frac{5}{7} \approx .714$$

as the rate of contraction, while considering

$$A^2 = \begin{pmatrix} \frac{19}{25} & \frac{27}{100} \\ \frac{6}{25} & \frac{73}{100} \end{pmatrix}$$

gives

$$s = \sqrt{\frac{27/100}{19/25} \cdot \frac{6/25}{73/100}} \approx .3418 \quad \Rightarrow \quad \Lambda \le \frac{.6582}{1.3418} \approx .4906 \approx (.7)^2,$$

a better estimate than we obtained from considering A itself.