Perturbation theory for Markov-type concentration of Markov chains

Elena Khusainova, Yale University

The goal of this technical note is to provide a review of perturbation theory results needed for the best available Markov chain concentration inequalities such as those by Gillman (1993), Gillman (1998), Lezaud (1998a), Paulin et al. (2015). The note is meant to be self-sufficient and all necessary proofs will be provided.

Perturbation theory studies the change of operator properties under the effect of small perturbations. In this note our main focus is on quantifying the effect of a certain type of perturbation (multiplying by a diagonal operator) on the transition operator spectrum, namely its largest eigenvalue. For more results from perturbation theory we refer the reader to Kato (2013).

We will start by reviewing complex analysis theory that constitutes the foundation of perturbation theory, continue with perturbation theory itself and conclude with the result on the form of the largest eigenvalue of a perturbed operator with respect to its Laurent series coefficients (will be defined in this section). In the appendix we illustrate how this theorem plays a pivotal role in proving concentration results for Markov chains with the Lezaud (1998a) concentration inequality for

$$Y_n := \sum_{t=0}^{n} f(X_t),$$

where $\{X_t,\ t = 0, 1, 2, ...\}$ is a Markov chain and $f : \mathbf{X} \to \mathbb{R}$ is a function on its state space $\mathbf{X}$.

1 Prerequisites: Complex analysis

In this section we provide the facts from complex analysis necessary for the development of perturbation theory. We will use the following notations:

- $C_r(z_0) = \{z \in \mathbb{C} : \ |z - z_0| = r\}$, a circle of radius r centered at z_0
- $D_r(z_0) = \{z \in \mathbb{C} : \ |z - z_0| < r\}$, an open disc of radius r centered at z_0
- $\overline{D}_r(z_0) = \{z \in \mathbb{C} : \ |z - z_0| \le r\}$, a closed disc of radius r centered at z_0
- $\mathring{D}_r(z_0) = \{z \in \mathbb{C} : \ 0 < |z - z_0| < r\}$, a punctured open disc of radius r centered at z_0

For the above we assume that $r > 0$.

Definition 1.1. *A sequence $\{z_n\}$ is a* **Cauchy sequence** *if for every $\varepsilon > 0$ there exists N such that $|z_m - z_n| \le \varepsilon$ for all $m, n > N$.*

Definition 1.2. *Consider an open set U, then a complex function $f(z)$ is* **holomorphic** *in U if* holomorphic

e-mail: elena.r.khusainova@gmail.com *version: 12 March 2020*

it is differentiable at every point in U, or in other words, the following limit exists

$$\lim_{h \to 0} \frac{f(z_0 + h) - f(z_0)}{h}$$

for any $z_0 \in U$.

Now we will define **Laurent series expansion** for a function through the following theorem.

Theorem 1.1. *Consider a complex function $f(z)$, holomorphic in punctured open disc $\mathring{D}_R(z_0)$. Let*

$$a_n := \frac{1}{2\pi i} \int\limits_{C_r(z_0)} \frac{f(z)}{(z - z_0)^{n+1}} dz,$$

where $0 < r < R$. Then

$$f(z) = \sum_{n=-\infty}^{+\infty} a_n (z - z_0)^n. \tag{1.1}$$

Laurent series

The right hand side of (1.1) is called Laurent series. It is also unique in a sense that if

$$f(z) = \sum_{n=-\infty}^{+\infty} a_n (z - z_0)^n = \sum_{n=-\infty}^{+\infty} b_n (z - z_0)^n,$$

then $a_n = b_n$.

residue

Definition 1.3. *The coefficient a_{-1} for the Laurent series of f is called* **residue of f at z_0**.

Remark 1.1. *Note that*

$$2\pi i a_{-1} = \int\limits_{C_r(z_0)} f(z) dz.$$

Definition 1.4. *We say that a parametrized curve $\gamma : [a, b] \to \mathbb{C}$ is* **smooth** *if $\gamma'(t)$ exists, is continuous on $[a, b]$ and $\gamma'(t) \neq 0$ for all $t \in [a, b]$. For the points $t = a$ and $t = b$ we interpret $\gamma'(t)$ as one-sided limits:*

$$\gamma'(a) = \lim_{h \to 0+} \frac{\gamma(a + h) - \gamma(a)}{h} \quad \text{and} \quad \gamma'(b) = \lim_{h \to 0-} \frac{\gamma(b + h) - \gamma(b)}{h}.$$

For a closed curve γ (that is $\gamma(a) = \gamma(b)$) we denote its open interior as $\mathcal{I}(\gamma)$.

Definition 1.5. *Let $\gamma : [a, b] \to \mathbb{C}$ be smooth and assuming that the range of γ belongs to the domain of f we define* **complex integral of f along** *γ (or* **around** *γ if γ is a closed curve) as*

$$\int\limits_{\gamma} f(z) dz := \int\limits_{a}^{b} f(\gamma(t)) \gamma'(t) dt.$$

Example 1.1. *If $\gamma(t) = e^{it}$ and $0 \leq t \leq 2\pi$ (that is γ is a unit circle centered at 0), then for an integer n:*

$$\int\limits_{\gamma} z^n dz = \begin{cases} 2\pi i & \text{if } n = -1 \\ 0 & \text{otherwise} \end{cases}$$

Proof. By definition

$$\begin{aligned}\int_\gamma z^n dz = \int_0^{2\pi} e^{nit} ie^{it} dt = \int_0^{2\pi} ie^{(n+1)it} dt = \\ =&\mathbb{1}\{n=-1\}\cdot i\int_0^{2\pi} dt + \mathbb{1}\{n\neq -1\}\cdot i\int_0^{2\pi} e^{(n+1)it}dt \\ =&\mathbb{1}\{n=-1\}\cdot 2\pi i + \mathbb{1}\{n\neq -1\}\cdot i\int_0^{2\pi} \cos\big((n+1)t\big) + i\sin\big((n+1)t\big)dt \\ =&\mathbb{1}\{n=-1\}\cdot 2\pi i + \mathbb{1}\{n\neq -1\}\cdot 0.\end{aligned}$$

□

Theorem 1.2. (Cauchy theorem) *Let $\gamma : [a, b] \to \mathbb{C}$ define a smooth closed curve and let function $f(z)$ be holomorphic in an open domain containing the closed interior of the curve, then*

$$\int_\gamma f(z)dz = 0.$$

Theorem 1.3. (Cauchy integral formula) *Let γ be a closed curve, and let f be holomorphic in an open domain containing the closed interior of the curve, then for every $a \in \mathcal{I}(\gamma)$:*

$$f(a) = \frac{1}{2\pi i}\int_\gamma \frac{f(z)}{z-a}dz.$$

Definition 1.6. *An isolated singularity $z = z_0$ of a function $f(z)$ with Laurent series*

$$f(z) = \sum_{n=-\infty}^{+\infty} a_n(z-z_0)^n$$

can be:

- **Essential singularity** *if $a_n \neq 0$ for infinitely many negative values of n. Otherwise define $\mathsf{ord}(f, z_0)$ to be the least such integer.*
- **Removable singularity** *if $\mathsf{ord}(f, z_0) \geq 0$ then we can eliminate the singularity by defining $f(z_0) := a_0$.*
- **Pole** *if $\mathsf{ord}(f, z_0) = m < 0$. For a pole type singularity we call the coefficient a_{-1} the* **residue** *of f at $z = z_0$.*

Theorem 1.4. (Residue theorem) *Consider a function $f(z)$ with holomorphic domain U containing the closed interior of the curve γ, except for finitely many poles c_i, $i = 1, ..., m$. Then*

$$\int_\gamma f(z)dz = 2\pi i\sum_{k=1}^{m} res_{z=c_k} f(z).$$

Theorem 1.5. (Integration by parts) *Let f and g be holomorphic functions on the domain U*

and consider curve $\gamma : [a, b] \to U$. *Then*

$$\int_\gamma f(z)g'(z)dz = -\int_\gamma f'(z)g(z)dz + \big[f(\gamma(b))g(\gamma(b)) - f(\gamma(a))g(\gamma(a))\big].$$

Note that for a closed curve the second term disappears.

Having now established the essential theory in complex analysis, we next examine the work of Kato (2013) and Lezaud (1998a) on the impact of perturbation of an operator on its spectrum. We are primarily interested in the largest eigenvalue of a transition matrix resulting from a specific perturbation: Multiplication by a "small" diagonal matrix. However, the theory works for any operator and perturbation.

2 Basic objects and their properties

In this section we define resolvent, perturbation and prove their basic properties. We assume that operators in question are finite.

resolvent

Definition 2.1. *For an operator* T *define its* ***resolvent*** *as*

$$\mathcal{R}(z) := (T - zI)^{-1}. \tag{2.1}$$

Properties and useful facts:

1° The resolvent has isolated singularities for $z \in \mathsf{spec}(T)$, where $\mathsf{spec}(T)$ is the spectrum of T.

2° Note that if $z_1, z_2 \notin \mathsf{spec}(T)$, then

$$\mathcal{R}(z_1) - \mathcal{R}(z_2) = (z_1 - z_2)\mathcal{R}(z_1)\mathcal{R}(z_2). \tag{2.2}$$

Proof. Follows from the chain of equivalences:

$$\begin{aligned}
\mathcal{R}(z_1) - \mathcal{R}(z_2) &= (z_1 - z_2)\mathcal{R}(z_1)\mathcal{R}(z_2)\\
\big(\mathcal{R}(z_1) - \mathcal{R}(z_2)\big)(T - z_2 I) &= (z_1 - z_2)\mathcal{R}(z_1)\mathcal{R}(z_2)(T - z_2 I)\\
\mathcal{R}(z_1)(T - z_2 I) - I &= (z_1 - z_2)\mathcal{R}(z_1)\\
(T - z_1 I)\mathcal{R}(z_1)(T - z_2 I) - (T - z_1 I) &= (z_1 - z_2)I\\
(T - z_2 I) - (T - z_1 I) &= (z_1 - z_2)I
\end{aligned}$$

□

3° If $z \notin \mathsf{spec}(T)$ then

$$T\mathcal{R}(z) = \mathcal{R}(z)T = I + z\mathcal{R}(z).$$

Proof.

$$T\mathcal{R}(z) - z\mathcal{R}(z) = (T - zI)\mathcal{R}(z) = I.$$

The second equality can be proved similarly. □

4° If v is an eigenvector of T, with corresponding eigenvalue λ, then it is also an eigenvector of $\mathcal{R}(z)$ for $z \neq \lambda$.

Proof. We have

$$\begin{aligned}\lambda v =& Tv\\ \lambda\mathcal{R}(z)v =& \mathcal{R}(z)Tv = v + z\mathcal{R}(z)v\\ \mathcal{R}(z)v =& \frac{1}{\lambda - z}v.\end{aligned}$$

□

Laurent series

Similar to complex functions we can write resolvent's Laurent series near the point $z_0 = \lambda$ as

$$\mathcal{R}(z) = \sum_{n=-\infty}^{\infty}(z-\lambda)^n A_n, \tag{2.3}$$

and define coefficients A_n as:

$$A_n = \frac{1}{2\pi i}\oint_{C_{r_n}(\lambda)}\frac{\mathcal{R}(z)}{(z-\lambda)^{n+1}}dz,$$

where $C_r(\lambda)$ is a circle centered at λ with radius r.

Remark 2.1. *Note that*

$$\begin{aligned}\sum_{n=-\infty}^{\infty}(z-\lambda)^n TA_n =& T\mathcal{R}(z) = I + z\mathcal{R}(z) = I + z\sum_{n=-\infty}^{\infty}(z-\lambda)^n A_n =\\ =& I + \sum_{n=-\infty}^{\infty}(z-\lambda)^{n+1}A_n + \lambda(z-\lambda)^n A_n =\\ =& \sum_{n=-\infty}^{\infty}(z-\lambda)^n(\lambda A_n + A_{n-1} + I\mathbb{1}_{n=0}).\end{aligned}$$

That is due to the uniqueness of Laurent series we have that

$$TA_n = \lambda A_n + A_{n-1} + I\mathbb{1}_{n=0},$$

or

$$(T-\lambda I)A_n = A_n(T-\lambda I) = A_{n-1} + I\mathbb{1}_{n=0}.$$

Theorem 2.1. *Consider a small punctured open disc $\mathring{D}_r(\lambda)$ centered at $\lambda \in \mathsf{spec}(T)$ with radius r small enough such that $\mathring{D}_r(\lambda) \cap \mathsf{spec}(T) = \lambda$ (such r exists as the spectrum of T is finite). For $z \in \mathring{D}_r(\lambda)$ we can rewrite the resolvent as* resolvent representation

$$\mathcal{R}(z) = \sum_{n=-\infty}^{\infty}(z-\lambda)^n A_n = -(z-\lambda)^{-1}H - \sum_{n=2}^{\infty}(z-\lambda)^{-n}U^{n-1} + \sum_{n=0}^{\infty}(z-\lambda)^n L^{n+1}, \tag{2.4}$$

where

$$\begin{aligned}H = -A_{-1} = -\frac{1}{2\pi i}\oint_{C_r(\lambda)}\mathcal{R}(z)dz,\\ U = -A_{-2} = -\frac{1}{2\pi i}\oint_{C_r(\lambda)}\mathcal{R}(z)(z-\lambda)dz,\end{aligned}$$

$$L = A_0 = \frac{1}{2\pi i} \oint\limits_{C_r(\lambda)} \frac{\mathcal{R}(z)}{(z-\lambda)} dz.$$

And also:

$$H^2 = H, \quad LH = HL = 0, \quad UH = HU = U.$$

Remark 2.2. *If we think of U, L and H as operators and denote $\mathbf{U}, \mathbf{L}, \mathbf{H}$ to be the corresponding ranges/images of the operators, then due to the last fact $\mathbf{U} \cap \mathbf{L} = \emptyset$, $\mathsf{dim}(\mathbf{U} \oplus \mathbf{L}) = \mathsf{rk}(\mathcal{R}(z))$ and H is a projection onto $\mathbf{U}$ along $\mathbf{L}$.*

Proof. (of theorem 2.1) We will start by proving the following:

$$A_n A_m = \big(\mathbb{1}_{n\geq 0} + \mathbb{1}_{m \geq 0} - 1\big) A_{n+m+1}.$$

First, note that:

$$\begin{aligned}
A_n A_m =& \frac{1}{2\pi i}\frac{1}{2\pi i} \oint\limits_{C_{r_m}(\lambda)} \oint\limits_{C_{r_n}(\lambda)} \frac{\mathcal{R}(z_n)}{(z_n-\lambda)^{n+1}} \frac{\mathcal{R}(z_m)}{(z_m-\lambda)^{m+1}} dz_n dz_m = \\
=& \frac{1}{(2\pi i)^2} \oint\limits_{C_{r_m}(\lambda)} \oint\limits_{C_{r_n}(\lambda)} \frac{\mathcal{R}(z_m) - \mathcal{R}(z_n)}{(z_m - z_n)(z_n-\lambda)^{n+1}(z_m-\lambda)^{m+1}} dz_n dz_m = \\
=& \frac{1}{2\pi i} \oint\limits_{C_{r_m}(\lambda)} \frac{\mathcal{R}(z_m)}{(z_m-\lambda)^{m+1}} \left[\frac{1}{2\pi i} \oint\limits_{C_{r_n}(\lambda)} \frac{1}{(z_m - z)(z-\lambda)^{n+1}} dz \right] dz_m + \\
&+ \frac{1}{2\pi i} \oint\limits_{C_{r_n}(\lambda)} \frac{\mathcal{R}(z_n)}{(z_n-\lambda)^{n+1}} \left[\frac{1}{2\pi i} \oint\limits_{C_{r_m}(\lambda)} \frac{1}{(z_n - z)(z-\lambda)^{m+1}} dz \right] dz_n.
\end{aligned} \tag{2.5}$$

Now, if we consider

$$h(z) := \frac{1}{(c-z)(z-\lambda)^{n+1}},$$

then the Laurent series for $h(z)$ around its singularities λ and c are the following:

$$\begin{aligned}
z_0 = \lambda: \ \ h(z) &= \sum_{s=-n-1}^{+\infty} \frac{1}{(c-z_0)^{s+n+2}} (z-z_0)^s, \\
z_0 = c: \ \ h(z) &= -\frac{1}{(z_0-\lambda)^{n+1}} \left[\sum_{s=0}^{+\infty} \left(\frac{z-z_0}{\lambda - z_0} \right)^s \right]^{n+1} \cdot \frac{1}{z-z_0}.
\end{aligned}$$

And using the residue theorem:

$$\frac{1}{2\pi i} \oint\limits_{C_r(\lambda)} \frac{1}{(c-z)(z-\lambda)^{n+1}} dz = \mathsf{res}_{z=\lambda} h(z) + \mathsf{res}_{z=c} h(z) \mathbb{1}\{c \in \overline{D}_r(\lambda)\} = \frac{\mathbb{1}\{n \geq 0\} - \mathbb{1}\{c \in \overline{D}_r(\lambda)\}}{(c-\lambda)^{n+1}}.$$

Assuming that $r_m > r_n$ we can then continue (2.5) as

$$\begin{aligned}
A_n A_m =& \frac{1}{2\pi i} \oint\limits_{C_{r_m}(\lambda)} \frac{\mathcal{R}(z_m)}{(z_m-\lambda)^{m+1}} \cdot \frac{\mathbb{1}\{n \geq 0\}}{(z_m-\lambda)^{n+1}} dz_m + \frac{1}{2\pi i} \oint\limits_{C_{r_n}(\lambda)} \frac{\mathcal{R}(z_n)}{(z_n-\lambda)^{n+1}} \cdot \frac{\mathbb{1}\{m \geq 0\} - 1}{(z_n-\lambda)^{m+1}} dz_n = \\
=& \frac{\mathbb{1}\{n \geq 0\}}{2\pi i} \oint\limits_{C_{r_n}(\lambda)} \frac{\mathcal{R}(z)}{(z-\lambda)^{n+m+1}} dz + \frac{\mathbb{1}\{m \geq 0\} - 1}{2\pi i} \oint\limits_{C_{r_m}(\lambda)} \frac{\mathcal{R}(z)}{(z-\lambda)^{n+m+1}} dz =
\end{aligned}$$

$$=\big(\mathbb{1}_{n\geq 0}+\mathbb{1}_{m\geq 0}-1\big)A_{n+m+1}.$$

Then

$$A_{-1}^2=-A_{-1},$$

thus $H=-A_{-1}$ is a projection. Also, $U^2=A_{-2}^2=-A_{-3}$, i.e $A_{-3}=-U^2$, $U^3=A_{-3}A_{-2}=-A_{-4}$, i.e $A_{-4}=-U^3$ and so on:

$$A_{-k}=-U^{k-1}.$$

Similarly, $A_1=A_0^2=L^2$, $A_2=A_1A_0=L^3$, ... , that is

$$A_k=L^{k+1}.$$

And finally,

$$LH=HL=-A_{-1}A_0=0 \text{ and } UH=HU=A_{-1}A_{-2}=-A_{-2}=U. \qquad \square$$

Remark 2.3. *If an eigenvalue λ has multiplicity m then the resolvent has pole of order m at $z_0=\lambda$ or equivalently, the coefficients A_{-n} in Laurent series are zero for $n>m$.*

For the rest of this note we will assume that operator T has an eigenvalue λ with multiplicity 1, which in the light of the above remark ensures that the resolvent doesn't have the "U-tail":

$$\mathcal{R}(z)=-(z-\lambda)^{-1}H+\sum_{n=0}^{\infty}(z-\lambda)^nL^{n+1}.$$

Before we proceed note that H is an eigenprojection:

Fact 2.1. *$HT=TH=\lambda H$, and therefore, H is a projection onto the space spanned by the corresponding eigenvector.*

Proof. Using that $HL=LH=0$ and

$$\mathcal{R}(z)=-(z-\lambda)^{-1}H+\sum_{n=0}^{\infty}(z-\lambda)^nL^{n+1},$$

we have

$$H\mathcal{R}(z)=\mathcal{R}(z)H=-(z-\lambda)^{-1}H.$$

Now, multiplying by $(T-zI)$ from the right we get

$$H=-(z-\lambda)^{-1}(T-zI)$$

or

$$(z-\lambda)H=zI-T.$$

If we multiply again from any side by H we get exactly what we need. To show that $\mathbf{U}$ is spanned by the eigenvector corresponding to λ, consider any other eigenvector v of T with eigenvalue $\lambda_v\neq\lambda$. Then

$$Hv=\frac{1}{\lambda}HTv=\frac{\lambda_v}{\lambda}Hv,$$

thus $Hv=0$. $\square$

perturbation

Definition 2.2. *Define the* **perturbation** *of* T

$$T(\theta) := T + \theta T_1 + \theta^2 T_2 + ... = T + T_\theta,$$

and its resolvent

$$\begin{aligned}\mathcal{R}(z,\theta) =& \big(T(\theta) - zI\big)^{-1} = \big(T - zI + T_\theta\big)^{-1} = \mathcal{R}(z)\big[I + T_\theta \mathcal{R}(z)\big]^{-1} = \\ =& \sum_{p=0}^{\infty} \mathcal{R}(z)\big(-T_\theta \mathcal{R}(z)\big)^p = \mathcal{R}(z) + \sum_{p=1}^{\infty} \mathcal{R}(z)\left[-\sum_{s=1}^{\infty} \theta^s T_s \mathcal{R}(z)\right]^p = \\ =& \mathcal{R}(z) + \sum_{p=1}^{\infty}(-1)^p \left[\sum_{s_1,\dots,s_p=1}^{\infty} \mathcal{R}(z)\theta^{s_1}T_{s_1}\mathcal{R}(z)\theta^{s_2}T_{s_2}\mathcal{R}(z)...\theta^{s_p}T_{s_p}\mathcal{R}(z)\right] = \\ =& \mathcal{R}(z) + \sum_{n=1}^{\infty}\theta^n \mathcal{R}_n(z),\end{aligned}$$

where

$$\mathcal{R}_n(z) = \sum_{p=1}^{n}(-1)^p \sum_{\substack{s_1+...+s_p=n \\ s_j\geq 1}} \mathcal{R}(z)T_{s_1}\mathcal{R}(z)T_{s_2}\mathcal{R}(z)...T_{s_p}\mathcal{R}(z).$$

Similar to before, we can also define operators $H(\theta)$, $U(\theta)$ and $L(\theta)$ for the resolvent $\mathcal{R}(z,\theta)$. Note though, that they depend on the singularity point at which the Laurent series is taken (we are not interested in the Laurent series around regular points). For the operator T the singularity points were its eigenvalues, which satisfy:

$$\mathsf{det}(T - \lambda I) = 0.$$

Due to the continuity for small enough θ the spectrum of $T(\theta)$ does not differ much from the one of T. Denote $\lambda(\theta)$ to be the largest eigenvalue of $T(\theta)$, and consider $H(\theta)$, $U(\theta)$ and $L(\theta)$ for $z = \lambda(\theta)$. Now we are ready for the main result.

3 Eigenvalue perturbation

eigenvalue perturbation

Theorem 3.1. *Let T be an operator with resolvent $\mathcal{R}(z)$ and largest eigenvalue λ. Then for its perturbation $T(\theta) = T + \sum_{s=1}^{\infty}\theta^n T_n$ with resolvent $\mathcal{R}(z,\theta)$ and largest eigenvalue $\lambda(\theta)$ define the coefficients λ_n such that:*

$$\lambda(\theta) = \lambda + \sum_{n=1}^{\infty}\theta^n \lambda_n.$$

Also recall that

$$L = \frac{1}{2\pi i}\oint_{C_r(\lambda)} \frac{\mathcal{R}(z)}{(z-\lambda)}dz, \qquad H = -\frac{1}{2\pi i}\oint_{C_r(\lambda)} \mathcal{R}(z)dz.$$

Then

$$\lambda_n = \sum_{p=1}^{\infty}\frac{(-1)^p}{p} \sum_{\substack{s_1+...+s_p=n \\ k_1+...+k_p=p-1 \\ s_j\geq 1,\ k_j\geq 0}} \mathsf{tr}\Big(T_{s_1}L^{(k_1)}T_{s_2}L^{(k_2)}...T_{s_p}L^{(k_p)}\Big), \tag{3.1}$$

where

$$L^{(k)} := L^k \mathbb{1}\{k > 0\} - H\mathbb{1}\{k = 0\}.$$

Before proceeding to the proof we need the following lemma:

Lemma 3.1.

$$\lambda(\theta) = \mathsf{tr}\big(T(\theta)H(\theta)\big) = \lambda + \mathsf{tr}\big((T(\theta) - \lambda I)H(\theta)\big).$$

Proof. First, recall that $H(\theta)$ is the projection onto the space spanned by the eigenvector corresponding to $\lambda(\theta)$, that is $\mathsf{rk}\big(H(\theta)\big) = \mathsf{tr}\big(H(\theta)\big) = 1$ and then

$$\begin{aligned}\lambda(\theta) =&\lambda(\theta)\mathsf{tr}\big(H(\theta)\big) = \mathsf{tr}\big(T(\theta)H(\theta)\big) = \mathsf{tr}\big(\lambda H(\theta) + (T(\theta) - \lambda I)H(\theta)\big) = \\ =&\lambda + \mathsf{tr}\big((T(\theta) - \lambda I)H(\theta)\big).\end{aligned}$$

□

Proof. (of theorem 3.1) In the proof we will be using several power series of θ, let's make a list of them to avoid confusion:

- $T(\theta)$. By definition:

$$T(\theta) = T + T_\theta = T + \sum_{n=1}^{\infty} \theta^n T_n.$$

- $\mathcal{R}(z, \theta)$. Also recall that

$$\mathcal{R}(z, \theta) = \sum_{p=0}^{\infty} \mathcal{R}\big(-T_\theta \mathcal{R}(z)\big)^p = \mathcal{R}(z) + \sum_{n=1}^{\infty} \theta^n \mathcal{R}_n(z),$$

  where

$$\mathcal{R}_n(z) = \sum_{p=1}^{n} (-1)^p \sum_{\substack{s_1+\ldots+s_p=n \\ s_j \geq 1}} \mathcal{R}(z)T_{s_1}\mathcal{R}(z)T_{s_2}\mathcal{R}(z)\ldots T_{s_p}\mathcal{R}(z).$$

- $H(\theta)$.

$$H(\theta) = -\frac{1}{2\pi i} \oint_{C_r(\lambda(\theta))} \mathcal{R}(z, \theta)dz = H + \sum_{n=1}^{\infty} \theta^n H_n(\theta),$$

  where

$$\begin{aligned}H_n =& -\frac{1}{2\pi i} \oint_{C_r(\lambda(\theta))} \mathcal{R}_n(z)dz = \\ =& -\frac{1}{2\pi i} \sum_{p=1}^{\infty} (-1)^p \sum_{\substack{s_1+\ldots+s_p=n \\ s_j \geq 1}} \oint_{C_r(\lambda(\theta))} \mathcal{R}(z)T_{s_1}\mathcal{R}(z)T_{s_2}\mathcal{R}(z)\ldots T_{s_p}\mathcal{R}(z)dz\end{aligned}$$

Using fact (3.1) and that $(T - \lambda I)H = 0$ (as λ has multiplicity 1) we get:

$$\begin{aligned}\lambda(\theta) =&\lambda + \mathsf{tr}\big((T(\theta) - \lambda I)H(\theta)\big) = \lambda + \mathsf{tr}\left[\Big(T - \lambda I + \sum_{n=1}^{\infty} \theta^n T_n\Big)H(\theta)\right] \\ =&\lambda + \mathsf{tr}\left[\Big(T - \lambda I + \sum_{n=1}^{\infty} \theta^n T_n\Big)\Big(H + \sum_{n=1}^{\infty} \theta^n H_n(\theta)\Big)\right]\end{aligned}$$

$$=\lambda + \sum_{n=1}^{\infty} \theta^n \underbrace{\mathsf{tr}\big[\widetilde{T}_n(\theta)\big]}_{\lambda_n}. \tag{3.2}$$

On the other hand:

$$\begin{aligned}
\lambda(\theta) =& \lambda + \mathsf{tr}\big[(T(\theta) - \lambda I)H(\theta)\big] \\
=& \lambda + \mathsf{tr}\Bigg[-\frac{1}{2\pi i}\oint_{C(\lambda(\theta),r)} \big(T(\theta) - \lambda I\big)\mathcal{R}(z,\theta)dz\Bigg] \\
=& \lambda + \mathsf{tr}\Bigg[-\frac{1}{2\pi i}\oint_{C(\lambda(\theta),r)} \big(T(\theta) - zI + zI - \lambda I\big)\mathcal{R}(z,\theta)dz\Bigg] = \\
=& \lambda + \mathsf{tr}\Bigg[-\frac{1}{2\pi i}\oint_{C(\lambda(\theta),r)} \big(z - \lambda\big)\mathcal{R}(z,\theta)dz\Bigg]. \qquad (3.3)
\end{aligned}$$

Then combining (3.2) with (3.3) we get:

$$\begin{aligned}
tr\Bigg[\sum_{n=1}^{\infty} \theta^n \widetilde{T}_n(\theta)\Bigg] =& -\frac{1}{2\pi i}\mathsf{tr}\Bigg[\oint_{C(\lambda(\theta),r)} (z-\lambda)\mathcal{R}(z,\theta)dz\Bigg] = \\
=& -\frac{1}{2\pi i}\mathsf{tr}\Bigg[\oint_{C(\lambda(\theta),r)} (z-\lambda)\Bigg(\mathcal{R}(z) + \sum_{p=1}^{\infty}\mathcal{R}(z)\big(-T_\theta\mathcal{R}(z)\big)^p\Bigg)dz\Bigg] = \\
=& -\frac{1}{2\pi i}\mathsf{tr}\Bigg[\oint_{C(\lambda(\theta),r)} (z-\lambda)\sum_{p=1}^{\infty}\mathcal{R}(z)\big(-T_\theta\mathcal{R}(z)\big)^p dz\Bigg], \qquad (3.4)
\end{aligned}$$

where the last equality follows from the fact that $(z-\lambda)\mathcal{R}(z)$ is holomorphic around $\lambda(\theta)$. Now note that

$$\frac{d}{dz}\mathcal{R}(z) = \mathcal{R}^2(z), \tag{3.5}$$

derivative trick

then for any operator $A(\theta)$ independent of z we have:

$$\frac{d}{dz}\big(A(\theta)\mathcal{R}(z)\big)^p = A\mathcal{R}^2A\mathcal{R}...\mathcal{R} + ... + A\mathcal{R}A\mathcal{R}...\mathcal{R}^2.$$

Then as $\mathsf{tr}(AB) = \mathsf{tr}(BA)$:

$$\mathsf{tr}\Bigg[\frac{d}{dz}\big(A(\theta)\mathcal{R}(z)\big)^p\Bigg] = p\ \mathsf{tr}\Big[\mathcal{R}(A\mathcal{R})^p\Big].$$

And thus we can continue (3.4) as

$$\begin{aligned}
tr\Bigg[\sum_{n=1}^{\infty} \theta^n \widetilde{T}_n(\theta)\Bigg] =& -\frac{1}{2\pi i}\mathsf{tr}\Bigg[\oint_{C(\lambda(\theta),r)} (z-\lambda)\sum_{p=1}^{\infty}\mathcal{R}(z)\big(-T_\theta\mathcal{R}(z)\big)^p dz\Bigg] = \\
=& -\frac{1}{2\pi i}\Bigg[\oint_{C(\lambda(\theta),r)} (z-\lambda)\sum_{p=1}^{\infty}(-1)^p\mathsf{tr}\Big[\mathcal{R}(z)\big(T_\theta\mathcal{R}(z)\big)^p\Big]dz\Bigg] =
\end{aligned}$$

$$= -\frac{1}{2\pi i}\mathsf{tr}\Bigg[\oint\limits_{C(\lambda(\theta),r)}(z-\lambda)\sum_{p=1}^{\infty}\frac{(-1)^p}{p}\Big[\frac{d}{dz}\big(T_\theta\mathcal{R}(z)\big)^p\Big]dz\Bigg].$$

Again, as there is no "U-tail", note that using integration by parts we get:

$$\begin{aligned}
\mathsf{tr}\Bigg[\sum_{n=1}^{\infty}\theta^n\widetilde{T}_n(\theta)\Bigg] &= -\frac{1}{2\pi i}\mathsf{tr}\Bigg[\oint\limits_{C(\lambda(\theta),r)}(z-\lambda)\sum_{p=1}^{\infty}\frac{(-1)^p}{p}\Big[\frac{d}{dz}\big(T_\theta\mathcal{R}(z)\big)^p\Big]dz\Bigg] = \\
&=\frac{1}{2\pi i}\mathsf{tr}\Bigg[\oint\limits_{C(\lambda(\theta),r)}\sum_{p=1}^{\infty}\frac{(-1)^p}{p}\big(T_\theta\mathcal{R}(z)\big)^p dz\Bigg] = \\
&=\frac{1}{2\pi i}\mathsf{tr}\Bigg[\oint\limits_{C(\lambda(\theta),r)}\sum_{p=1}^{\infty}\frac{(-1)^p}{p}\sum_{k_1+\ldots+k_p}(z-\lambda)^{k_1+..+k_p-p}T_\theta L^{(k_1)}T_\theta\ldots T_\theta L^{(k_p)}dz\Bigg] = \\
&=\mathsf{tr}\sum_{p=1}^{\infty}\frac{(-1)^p}{p}\sum_{\substack{k_1+\ldots+k_p=p-1\\ k_j\ge 0}}T_\theta L^{(k_1)}T_\theta\ldots T_\theta L^{(k_p)}. \qquad (3.6)
\end{aligned}$$

or

$$\lambda_n = \mathsf{tr}(\widetilde{T}^{(n)}(\theta)) = \sum_{p=1}^{\infty}\frac{(-1)^p}{p}\sum_{\substack{s_1+\ldots+s_p=n\\ k_1+\ldots+k_p=p-1\\ s_j\ge 1,\ k_j\ge 0}}\mathsf{tr}\Big[T_{s_1}L^{(k_1)}T_{s_2}L^{(k_2)}\ldots T_{s_p}L^{(k_p)}\Big], \qquad (3.7)$$

which completes the proof. □

Remark 3.1. *The derivative trick (3.5) can be avoided using only the invariance of a trace to rotations:*

$$\begin{aligned}
\mathsf{tr}\Bigg[\sum_{n=1}^{\infty}\theta^n\widetilde{T}_n(\theta)\Bigg] &= -\frac{1}{2\pi i}\mathsf{tr}\Bigg[\oint\limits_{C(\lambda(\theta),r)}(z-\lambda)\sum_{p=1}^{\infty}\mathcal{R}(z)\big(-T_\theta\mathcal{R}(z)\big)^p dz\Bigg] = \\
&= -\frac{1}{2\pi i}\mathsf{tr}\Bigg[\oint\limits_{C(\lambda(\theta),r)}(z-\lambda)\sum_{p=1}^{\infty}(-1)^p\sum_{k_0=0}^{\infty}(z-\lambda)^{k_0-1}L^{(k_0)}\times \\
&\qquad\qquad \times\sum_{k_1,\ldots,k_p}(z-\lambda)^{k_1+\ldots+k_p-p}T_\theta L^{(k_1)}\ldots L^{(k_p)}dz\Bigg] = \\
&= -\mathsf{tr}\sum_{p=1}^{\infty}(-1)^p\sum_{\substack{k_0+k_1+\ldots+k_p=p-1\\ k_j\ge 0}}T_\theta L^{(k_1)}\ldots L^{(k_p)}L^{(k_0)}
\end{aligned}$$

Now let us take a closer look at the term $L^{(k_p)}L^{(k_0)}$. From the definition of $L^{(k)}$ we know that:

$$\begin{aligned}
&L^{(0)}L^{(0)} = (-H)(-H) = H^2 = H = -L^{(0)},\\
&L^{(0)}L^{(k)} = L^{(k)}L^{(0)} = 0,\ \textit{if } k>0,\\
&L^{(k_1)}L^{(k_2)} = L^{(k_1+k_2)},\ \textit{if } k_1>0, k_2>0.
\end{aligned}$$

Thus:

$$\operatorname{tr}\left[\sum_{n=1}^{\infty}\theta^n\widetilde{T}_n(\theta)\right] = -\operatorname{tr}\sum_{p=1}^{\infty}(-1)^p \sum_{\substack{k_0+k_1+\ldots+k_p=p-1\\ k_j\geq 0}} T_\theta L^{(k_1)}\ldots L^{(k_p)}L^{(k_0)}$$

$$= -\operatorname{tr}\sum_{p=1}^{\infty}(-1)^p \sum_{k_1=0}^{p-1} T_\theta L^{(k_1)} \sum_{k_2=0}^{p-1} T_\theta L^{(k_2)}\ldots \sum_{k_{p-1}=0}^{p-1} T_\theta L^{(k_{p-1})}\times$$

$$\times \sum_{k_0,k_p=0}^{p-1} T_\theta L^{(k_p)}L^{(k_0)}\mathbb{1}\{k_0+\ldots+k_p=p-1\}$$

$$= -\operatorname{tr}\sum_{p=1}^{\infty}(-1)^p \sum_{k_1=0}^{p-1} T_\theta L^{(k_1)}\ldots \sum_{k_{p-1}=0}^{p-1} T_\theta L^{(k_{p-1})} \sum_{k_p=0}^{p-1}(k_p-1)T_\theta L^{(k_p)}\mathbb{1}\{k_1+\ldots+k_p=p-1\} =$$

$$= -\operatorname{tr}\sum_{p=1}^{\infty}(-1)^p \sum_{k_1+\ldots+k_p=p-1, k_j\geq 0}(k_p-1)T_\theta L^{(k_1)}\ldots T_\theta L^{(k_p)} =$$

$$= -\operatorname{tr}\sum_{p=1}^{\infty}(-1)^p \sum_{\max_{lex}\left\{[k_1,\ldots,k_p]:\ k_1+\ldots+k_p=p-1,\ k_j\geq 0\right\}} T_\theta L^{(k_1)}T_\theta\ldots T_\theta L^{(k_p)}\sum_{j=1}^{p}(k_j-1) =$$

$$= \operatorname{tr}\sum_{p=1}^{\infty}(-1)^p \sum_{\max_{lex}\left\{[k_1,\ldots,k_p]:\ k_1+\ldots+k_p=p-1,\ k_j\geq 0\right\}} T_\theta L^{(k_1)}T_\theta\ldots T_\theta L^{(k_p)}.$$

Note that as the trace is invariant to the rotations of the coefficients $k_1,..,k_p$, we can pick one (let's say lexicographically maximal) and then multiply the result by p (number of rotations). Then

$$(3.6) = \operatorname{tr}\sum_{p=1}^{\infty}(-1)^p \sum_{\max_{lex}\left\{[k_1,\ldots,k_p]:\ k_1+\ldots+k_p=p-1,\ k_j\geq 0\right\}} T_\theta L^{(k_1)}T_\theta\ldots T_\theta L^{(k_p)},$$

resulting in the same expression.

Note that the above theorem holds in general case, but for applying it in Markov chain context, we only need to assume a certain type of perturbation. This type of perturbation can be expressed as multiplication by another operator: $T(\theta) = TD$, where operator D is diagonal with "small" entries. The consequences of such perturbation are illustrated by the following example.

transition operator
diagonal perturbation

Example 3.1. Perturbation of a transition operator. *Consider a transition operator P of an irreducible Markov chain with stationary distribution π and vector g such that $\pi^T g = 0$. Also, define a diagonal operators G and E_r such that $G = diag(g)$ and*

$$E_r[i,j] = \begin{cases} e^{rg[i]} & \text{if } i=j \\ 0 & \text{otherwise.}\end{cases}$$

Consider the following perturbation $P(r)$ of operator P:

$$P(r) = PE_r = P + rPG + \frac{r^2}{2!}PG^2 + \ldots$$

Then if $\lambda(r)$ is the largest eigenvalue of P_r, then

$$\lambda(r) = \lambda + \sum_{n=1}^{\infty} r^n \lambda_n.$$

where

$$\lambda_n = \sum_{p=1}^{n} \frac{(-1)^p}{p} \sum_{\substack{s_1+\ldots+s_p=n \\ k_1+\ldots+k_p=p-1 \\ s_j\geq 1,\ k_j\geq 0}} \frac{1}{s_1!\ldots s_p!} \langle g^{s_1}, L^{(k_1)} P G^{s_2} \ldots L^{(k_{p-1})} P g^{s_p} \rangle_\pi \tag{3.8}$$

with

$$L^{(0)} = -H = -\mathbf{1}\pi^T, \qquad L^{(k)} = L^k, \text{ for } k \geq 1,$$

where $\mathbf{1}$ *denotes a vector of ones.*

Before proceeding to the proof let us make the following remark.

Remark 3.2. *If $k_1 + \ldots + k_p = p - 1$ and $k_j \geq 0$ then an least one of k_i has to be 0 (w.l.o.g. let $k_p = 0$), then using the fact that trace is invariant under cyclic permutations of its arguments we get*

$$\mathsf{tr}\Big[T_{s_1} L^{(k_1)} \ldots T_{s_p} L^{(k_p)}\Big] = -\,\mathsf{tr}\Big[T_{s_1} L^{(k_1)} \ldots T_{s_p} H\Big] = -\mathsf{tr}\Big[H T_{s_1} L^{(k_1)} \ldots T_{s_p}\Big].$$

Additionally, if $G = diag(g)$ is a diagonal matrix and π is a vector of weights, then for any matrix M:

$$\begin{aligned}\mathsf{tr}\big[\mathbf{1}\pi^T G^{k_1} M G^{k_2}\big] &= \sum_{i=1}^{n} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \sum_{j_3=1}^{n} \pi[j_1] G^{k_1}[j_1, j_2] M[j_2, j_3] G^{k_2}[j_3, i] = \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \pi_j g_j^{k_1} M[j, i] g_i^{k_2} = \langle g^{k_1}, M g^{k_2} \rangle_\pi.\end{aligned}$$

Proof. (of example 3.1) We will start by showing that the eigenprojector H for operator P corresponding to eigenvalue 1 has the following form:

$$H = \mathbf{1}\pi^T.$$

Indeed, H is a projection as

$$H^2 = \mathbf{1}\pi^T \mathbf{1}\pi^T = H,$$

and for v such that $\pi^T v = 0$

$$Hv = \mathbf{1}\pi^T v = 0.$$

Applying (3.7) we get

$$\lambda_n = \sum_{p=1}^{n} \frac{(-1)^p}{p} \sum_{\substack{s_1+\ldots+s_p=n \\ k_1+\ldots+k_p=p-1 \\ s_j\geq 1,\ k_j\geq 0}} \frac{1}{s_1!\ldots s_p!} \mathsf{tr}\Big[P G^{s_1} L^{(k_1)} \ldots P G^{s_p} L^{(k_p)}\Big]. \tag{3.9}$$

Due to remark 3.2 we have

$$\begin{aligned}\mathsf{tr}\Big[PG^{s_1}L^{(k_1)}...PG^{s_p}L^{(k_p)}\Big] =&\mathsf{tr}\Big[L^{(k_p)}PG^{s_1}L^{(k_1)}...PG^{s_p}\Big] = \\ =&\mathsf{tr}\Big[\mathbf{1}(\pi^T P)G^{s_1}L^{(k_1)}PG^{s_2}...L^{(k_{p-1})}PG^{s_p}\Big] = \\ =&\langle g^{s_1}, L^{(k_1)}PG^{s_2}...L^{(k_{p-1})}Pg^{s_p}\rangle_\pi .\end{aligned}$$

Then we can rewrite (3.9) as

$$\lambda_n = \sum_{p=1}^{n} \frac{(-1)^p}{p} \sum_{\substack{s_1+...+s_p=n \\ k_1+...+k_p=p-1 \\ s_j\geq 1,\ k_j\geq 0}} \frac{1}{s_1!...s_p!}\langle g^{s_1}, L^{(k_1)}PG^{s_2}...L^{(k_{p-1})}Pg^{s_p}\rangle_\pi . \tag{3.10}$$

And, for instance, $\lambda_1 = -\langle g, \mathbf{1}\rangle_\pi = 0$. The example is proved. $\square$

Bibliography

Gillman, D. (1998). A chernoff bound for random walks on expander graphs. *SIAM Journal on Computing 27*(4), 1203–1220.

Gillman, D. W. (1993). *Hidden Markov chains: convergence rates and the complexity of inference.* Ph. D. thesis, Massachusetts Institute of Technology.

Kato, T. (2013). *Perturbation theory for linear operators*, Volume 132. Springer Science & Business Media.

Lezaud, P. (1998a). Chernoff-type bound for finite markov chains. *Annals of Applied Probability*, 849–867.

Lezaud, P. (1998b). *Étude quantitative des chaînes de Markov par perturbation de leur noyau.* Ph. D. thesis, Université Toulouse 3 Paul Sabatier.

Paulin, D. et al. (2015). Concentration inequalities for markov chains by marton couplings and spectral methods. *Electronic Journal of Probability 20.*

Appendix

In this section we present one of many applications of perturbation theory in the field of Markov chains. The paper by Lezaud (1998a) was published in an attempt to summarize and translate the author's PhD thesis (Lezaud (1998b), in French). Presumable, due to the length restrictions the proof of the result skips important steps and is generally difficult to follow. This section together with the perturbation theory results covered above constitute the full proof. Lezaud (1998a)

Theorem 3.2. *Let P be a transition matrix of an irreducible and reversible Markov chain $\{X_1, X_2, ...\}$ on a finite set $\mathbf{X}$, having a stationary distribution π. Let $g : \mathbf{X} \to \mathbb{R}$ be such that $\mathbb{E}_\pi(g) = \pi^T g = 0$, $\|g\|_\infty \le 1$ and $0 < \mathbb{E}_\pi(g)^2 \le b^2$. Then for any positive integer n and all $0 < \varepsilon \le 1$, assuming that chain starts at distribution π_0 and $\varepsilon \le 2b^2/5$:*

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^n g(X_i) \ge \varepsilon\right] \le \exp\left(\frac{\gamma}{5} - \frac{n\varepsilon^2\gamma\cdot(1-5\varepsilon/2b^2)}{4b^2}\right),$$

where γ is the spectral gap of P.

Proof. Define

$$Y_n := \sum_{j=1}^n g(X_j).$$

Define E_r and P_r similar to example 3.1 and let $F_r := \sqrt{E_r}$, and $S := F_r P F_r$. Then S is similar to P_r (in a sense that is has same eigenvalues) and the following relations are true:

$$\begin{aligned} P_r &= PE_r = F_r^{-1}F_rPF_rF_r = F_r^{-1}SF_r, \\ P_r^n &= \underbrace{(F_r^{-1}SF_r)...(F_r^{-1}SF_r)}_{n\ times} = F_r^{-1}S^nF_r. \end{aligned}$$

Note that $\mathbb{E}_{\pi_0}\exp(rY_n) = \pi_0^T P_r^n \mathbf{1}$ and then applying Markov inequality we get

$$\begin{aligned} \mathbb{P}_{\pi_0}\Big\{Y_n \ge n\varepsilon\Big\} =& \mathbb{P}_{\pi_0}\Big\{rY_n \ge rn\varepsilon\Big\} \le e^{-rn\varepsilon}\mathbb{E}_{\pi_0}\exp(rY_n) \\ =& e^{-rn\varepsilon}\pi_0^T P_r^n\mathbf{1} = e^{-rn\varepsilon}\pi_0^T F_r^{-1}S^nF_r\mathbf{1}, \end{aligned} \tag{3.11}$$

where $r > 0$. That is, we need to upper bound $\pi_0^T F_r^{-1} S^n F_r \mathbf{1}$. Note that S is self-adjoint w.r.t. π:

$$\langle x, Sy\rangle_\pi = \langle Sx, y\rangle_\pi \text{ for all } x, y.$$

And thus $\|S\|_{l_2(\pi)\to l_2(\pi)} = \lambda_0(S) = \lambda_0(P_r)$, where λ_0 is the largest eigenvalue. Then

$$\begin{aligned} \pi_0^T F_r^{-1}S^nF_r\mathbf{1} =& \Big\langle \frac{\pi_0}{\pi}, F_r^{-1}S^nF_r\mathbf{1}\Big\rangle_\pi \le \Big\|\frac{\pi_0}{\pi}\Big\|_\pi \big\|F_r^{-1}S^nF_r\mathbf{1}\big\|_\pi \le \\ \le& \Big\|\frac{\pi_0}{\pi}\Big\|_\pi \big\|F_r^{-1}\big\|_\pi \big\|S^n\big\|_\pi \big\|F_r\mathbf{1}\big\|_\pi \le N_{\pi_0}e^r\lambda_0^n(P_r), \end{aligned}$$

where with some abuse of notation $\frac{\pi_0}{\pi}$ denotes resulting vector of the element-wise division and $N_{\pi_0} := \big\|\frac{\pi_0}{\pi}\big\|_\pi$. Then we can continue (3.11) as

$$\mathbb{P}_{\pi_0}\Big\{Y_n \ge n\varepsilon\Big\} \le e^{-rn\varepsilon}N_{\pi_0}e^r\lambda_0^n(P_r) = e^rN_{\pi_0}\exp\Big(-n\big(r\varepsilon - \log\lambda_0(P_r)\big)\Big). \tag{3.12}$$

Now in the light of the example 3.1 we have that

$$\lambda_0(P_r) = 1 + \sum_{n=2}^{\infty} r^n \lambda_n,$$

where

$$\lambda_n = \sum_{p=1}^{n} \frac{(-1)^p}{p} \sum_{\substack{s_1+...+s_p=n \\ k_1+...+k_p=p-1 \\ s_j \geq 1,\ k_j \geq 0}} \frac{1}{s_1!...s_p!} \langle g^{s_1}, L^{(k_1)} P G^{s_2} ... L^{(k_{p-1})} P g^{s_p} \rangle_\pi, \tag{3.13}$$

with

$$L^{(0)} = -H = -\mathbf{1}\pi^T, \qquad L^{(k)} = L^k \text{ for } k \geq 0.$$

Note that we can bound each term using Cauchy-Schwarz together with the fact that $\|g\|_\infty \leq 1$ and $\|g\|_\pi^2 \leq b$:

$$\begin{aligned} \langle g^{s_1}, L^{(k_1)} P G^{s_2} ... L^{(k_{p-1})} P g^{s_p} \rangle_\pi \leq & \|g^{s_1}\|_\pi \|g^{s_p}\|_\pi \|G\|_\pi^{n-s_1-s_p} \|L\|_\pi^{p-1} \|P\|_\pi^{p-1} \\ \leq & b^2 \|G\|_\pi^{n-s_1-s_p} \|L\|_\pi^{p-1} \|P\|_\pi^{p-1} \\ \leq & \frac{b^2}{\gamma^{n-1}}. \end{aligned} \tag{3.14}$$

The number of terms in (3.13) can be bound as:

$$C(n) = \binom{n-1}{p-1} \binom{2(p-1)}{p-1} \leq \frac{5^n}{4n} \leq 5^{n-2}, \quad \text{if } n \geq 7,$$

and combining with (3.14) we get:

$$\lambda_n \leq \frac{b^2}{\gamma^{n-1}} \cdot 5^{n-2} = \frac{b^2}{\gamma} \left(\frac{5}{\gamma} \right)^{n-2}, \quad n \geq 2.$$

Thus

$$\lambda_0(P_r) = 1 + r^2 \sum_{n=0}^{\infty} r^n \lambda_{n+2} \leq 1 + \frac{b^2}{\gamma} r^2 \sum_{n=0}^{\infty} \left(\frac{5r}{\gamma} \right)^n = 1 + \frac{b^2}{\gamma} r^2 \left(1 - \frac{5r}{\gamma} \right)^{-1}.$$

And using the fact that $\log(1+x) \leq x$ we can continue (3.12) as

$$\begin{aligned} \mathbb{P}_{\pi_0} \Big\{ Y_n \geq n\varepsilon \Big\} \leq & e^r N_{\pi_0} \exp\Big(-n \big(r\varepsilon - \log \lambda_0(P_r) \big) \Big) \\ \leq & e^r N_{\pi_0} \exp\left(-n \left(r\varepsilon - r^2 \frac{b^2}{\gamma} \Big(1 - \frac{5r}{\gamma} \Big)^{-1} \right) \right). \end{aligned} \tag{3.15}$$

We are free in the choice of $r > 0$, and we will choose it to maximize function $Q(r)$, where

$$Q(r) = r\varepsilon - r^2 \frac{b^2}{\gamma} \Big(1 - \frac{5r}{\gamma} \Big)^{-1},$$

which means

$$r = \frac{\varepsilon\gamma}{b^2 \left(1 + 5\varepsilon/b^2 + \sqrt{1 + 5\varepsilon/b^2} \right)}.$$

Plugging in this r in (3.15) will result in the desired bound. $\square$