Theorem (FT of MC)

- Consider $b_t = a_t P - a_t$ (Net gain vector after one step)

$$b_t = \frac{1}{t}[p_0 P + p_1 P + \dots + p_{t-1} P] - \frac{1}{t}[p_0 + p_1 + \dots + p_{t-1}]$$

$$= \frac{1}{t}[p_1 + \dots + p_t] - \frac{1}{t}[p_0 + \dots + p_{t-1}]$$

(P is the one-step transition)

$$= \frac{1}{t}[p_t - p_0]$$

<u>Proposition</u>: For two probability distributions p & q

$$\| p - q \|_1 = 2 \sum_i (p_i - q_i)^+ = 2 \sum_i (q_i - p_i)^+$$

Proof:

$$\begin{cases} \| p - q \|_1 = \sum_{p_i > q_i} (p_i - q_i) + \sum_{p_i < q_i} (q_i - p_i) & \dots ① \\ 0 = \sum_{p_i > q_i} (p_i - q_i) - \sum_{p_i < q_i} (q_i - p_i) & \dots ② \end{cases}$$

Compute ① + ② , ① − ②, proposition proved. ☒

Using proposition, $\| b_t \|_1 = \frac{2}{t} \sum_i (p_t - p_0)_i^+ \leq \frac{2}{t}$

<u>Lemma</u>: The $n \times (n+1)$ matrix $A = [P - I, \mathbb{1}_n]$ has rank n if P is the transition matrix for a connected Markov chain.

Proof: Use contradiction.

- Suppose rank $(A) < n$, then there would be at least 2 linearly independent solutions to $Ax = 0$
  (Rank-nullity: rank (A) + nullity$(A) = n+1$)

- Since P is stochastic (row sum = 1), $(\bar{1}_n, 0)^T$ is a solution to $Ax = 0$.

- Suppose $(\bar{x}_n, \alpha)^T$ is a solution orthogonal to $(\bar{1}_n, 0)^T$

we have: $x_i = \sum_j p_{ij} x_j + \alpha$ -- ③

- Let $S = \{ i : x_i = \max_{j=1}^{n} x_j \}$, observe that:
  1) First, $\bar{x}_n \neq \bar{0}_n$, since if otherwise, $\alpha = 0$ by ③, then $(\bar{x}_n, \alpha)^T = \bar{0}_{n+1}$
  2) Then by orthogonality, $\bar{S}$ can't be empty.

- Since the Markov chain is connected. $\exists (k, l) \in (S, \bar{S})$ s.t. $p_{kl} > 0$, $\Rightarrow a > 0$ by ③

- However, consider symmetric argument using $S' = \{ i : x_i = \max_{j=1}^{n} x_j \}$, one can show $a < 0$, ※ ☒

- By lemma, $A = [P-I, \bar{1}_n]$ has rank n.

  Note that first n columns sum to 0, thus linearly dependent.

- Consider $B = A_{1:n, 2:(n+1)}$, i.e., B is obtained by removing 1^{st} column of A.

  Observe: B is clearly invertible.

- Define the $(n-1)$- vector c_t^T by $a_t B = (c_t, 1)$

  Alternatively, $c_t = (b_t)_{2:n}$

- Since $\|b_t\|_1 \leq \frac{2}{t}$, $b_t \to \bar{0}_n$. i.e., $c_t \to \bar{0}_{n-1}$

  $\therefore a_t = (c_t, 1) B^{-1} \to (\bar{0}_{n-1}, 1) B^{-1} \triangleq \pi$ "Define to be."

- Since a_t is a probability vector, $\lim_{t\to\infty} a_t = \pi$ is as well.

  (Inner product is a linear function, can interchange order of function and limit)

- Since $\lim_{t\to\infty} (a_t (P-I)) = 0 \Rightarrow (\lim_{t\to\infty} a_t)(P-I) = 0$

  i.e $\pi = \pi P$

- The uniqueness is guaranteed by $\text{rank}(A) = n$ and π is a solution to $A^T x = (\bar{0}_n, 1)^T$ ($A^T : (n+1) \times n$)

Theorem 5.5

- Let $t = \frac{c \ln(1/\pi_{min})}{\Phi^2 \varepsilon^3}$ for suitable constant c
- Let $a = a_t = \frac{1}{t}[p_1 + p_2 + \dots + p_t]$
- Let $v_i = \frac{\pi_i}{a_i}$ and renumber the states s.t v_i's are in descending order:

$$v_1 \geq v_2 \geq \dots v_{i_0} \geq 1 > v_{i_0+1} \geq \dots \geq v_n$$

Goal: Show $\|a - \pi\|_1 \leq \varepsilon$

Roadmap:

1) Reduce the problem to proving v_i's do not drop too fast as we increase i

2) Prove the reduced problem stated above.

\# Intuition: Consider the extreme case when $a = \pi$, all v_i equal 1. If $a \to \pi$, v should approach 1_n, so our argument makes sense.

- We call a state i <u>heavy</u> if $v_i > 1$.

  It has more probability mass according to a than its stationary distribution.

- Let i_0 be the last heavy state, i.e

$$i_0 = \max_{i: v_i > 1} i$$

we have

$$\| a - \pi \|_1 = \sum_{i: a_i > \pi_i} a_i - \pi_i + \sum_{i: \pi_i > a_i} \pi_i - a_i$$

a, π both sum to 1.

$$= 2 \sum_{i=1}^{i_0} (v_i - 1) \pi_i$$

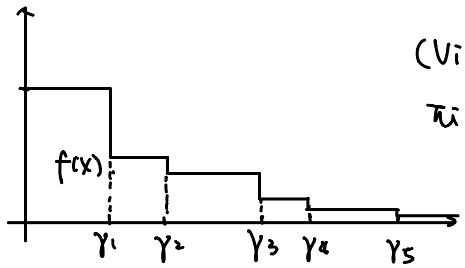
$$= 2 \sum_{i=i_0+1}^{n} (1 - v_i) \pi_i$$

- let $\gamma_i = \pi_1 + \pi_2 + \cdots + \pi_i$ for $1 \leq i \leq n$, $\gamma_0 = 0$

  Define a piecewise-constant (step) function $f : [0, \gamma_{i_0}] \to \mathbb{R}$ by $f(x) = v_i - 1$ for $x \in [\gamma_{i-1}, \gamma_i)$

  Now

$$\frac{\| a - \pi \|_1}{2} = \sum_{i=1}^{i_0} (v_i - 1) \pi_i = \int_0^{\gamma_{i_0}} f(x) \, dx$$



$(v_i - 1)$: height of the rectangles

π_i : width of the rectangles.

- We now give an upper bound of $\int_0^{\gamma_{i_0}} f(x)\,dx$

We partition the heavy states $\{1, 2, \dots, i_0\}$ into r contiguous groups: $G_1, \dots, G_r$.

i.e. $(1 + \text{largest element of } G_k) = (\text{least element of } G_{k+1})$

※ We will specify the groups later.

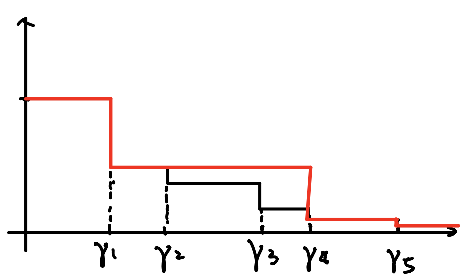
- Let $u_k = \max_{i \in G_k} v_i$ and $u_{r+1} = 0$

We define a function $g : [0, \gamma_{i_0}] \to \mathbb{R}$ by

$g(x) := u_k$ for $x \in \bigcup_{i \in G_k} [\gamma_{i-1}, \gamma_i)$

Since $f(x) \le g(x)$ for all x,

$\int_0^{\gamma_{i_0}} f(x)\,dx \le \int_0^{\gamma_{i_0}} g(x)\,dx$

- 

Suppose $G_1 = \{1\}$
$G_2 = \{2, 3, 4\}$
$G_3 = \{5\}$
...

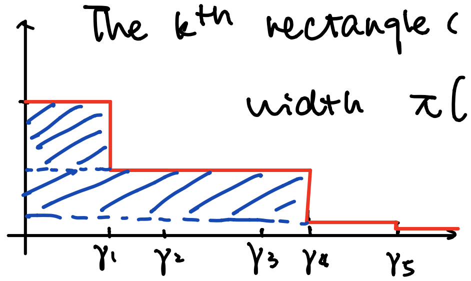
Then $g(x)$ is shown by

- Claim:

$$\int_0^{\gamma_{i_0}} g(x) = \sum_{k=1}^{r} \pi(G_1 \cup G_2 \cup \dots \cup G_k)(u_k - u_{k+1})$$

(Informal Proof)

Divide the area under the curve into vertically stacked rectangles.



The k^{th} rectangle (Top → Bottom) has height $u_k - u_{k+1}$, width $\pi(G_1 \cup G_2 \cup \dots \cup G_k)$

- Now it suffices to show

$$\sum_{k=1}^{r} \pi(G_1 \cup G_2 \cup \dots \cup G_k)(u_k - u_{k+1}) \le \frac{\varepsilon}{2} \quad \dots \quad (*)$$

- We observe that

If $\sum_{i=i_0+1}^{n} \pi_i \le \frac{\varepsilon}{2}$, we are done since:

$$\|a - \pi\|_1 = 2 \sum_{i=i_0+1}^{n} (1 - \underbrace{v_i}_{\text{Nonnegative, hence } 1-v_i \le 1}) \pi_i \le 2 \sum_{i=i_0+1}^{n} \pi_i \le \varepsilon$$

So we consider $\sum_{i=i_0+1}^{n} \pi_i > \frac{\varepsilon}{2} \Rightarrow \varepsilon < 2\sum_{i=i_0+1}^{n} \pi_i < 2$

For any subset A of heavy states, we have

$$\min(\pi(A), \pi(\bar{A})) \geq \frac{\varepsilon}{2}\pi(A) \quad \cdots (**)$$

- $\varepsilon < 2$, $\therefore \pi(A) > \frac{\varepsilon}{2}\pi(A)$.
- $\pi(\bar{A}) > \sum_{i=i_0+1}^{n} \pi_i > \frac{\varepsilon}{2} > \frac{\varepsilon}{2}\cdot\pi(A)$

- Now define the subsets in a recursive manner.

→ $G_1 = \{1\}$

→ Suppose $G_1, \dots, G_{k-1}$ are defined and that largest element of G_{k-1} is $(s-1)$. so that G_k starts from s.

→ Let l, the last element of G_k to be the

▹ Largest element $\geq s$ and $\leq i_0$ such that

$$\sum_{j=s+1}^{l} \pi_j \leq \frac{\varepsilon \Phi \gamma_s}{4} \quad \cdots (***)$$

↳ So when this not satisfied for any $l \geq s+1$, $l = s$, i.e. $G_k = \{s\}$.

· Now we prove a lemma:

Lemma: Suppose G, u defined as above

$$\pi(G_1 \cup G_2 \cup \dots \cup G_k)(u_k - u_{k+1}) \le \frac{8}{t\Phi\varepsilon} + \frac{2}{t}$$

Remark: This lemma together with an upper bound of r is going to lead to $(*)$.

· Idea: Using 2 ways to calculate "probability flow" from heavy states to light states. Starting at probability vector a. It transitions to aP. So the net loss of probability for state j due to the step is $[a_j - (aP)_j]$

Consider a particular group $G_k = \{s, s+1, \dots, l\}$

Suppose $s < i_0$. Let $A = \{1, \dots, s\}$.

① (By proof of FT of MC) The net loss of probability is

$$\sum_{i=1}^{s} (a_i - (aP)_i) \le \sum_{i=1}^{n} |a_i - (aP)_i| \le \frac{2}{t}$$

② (By taking difference of flow from $A \to \bar{A}$ and from $\bar{A}$ to A)

For any state i, j with $i < j$:

$$\text{net-flow}(i, j) = \text{flow}(i, j) - \text{flow}(j, i)$$

$$= \underbrace{v_i \pi_i}_{a_i} p_{ij} - \underbrace{v_j \pi_j}_{a_j} p_{ji}$$

$$= (v_i - v_j)\, \pi_j p_{ji} \quad (\text{Detailed balance})$$

$$\geq 0$$

Thus for states $i < j$, $\text{net-flow}(i, j) \geq 0$

Hence, $$\frac{2}{t} \geq \text{net-flow}(A, \bar{A})$$

$$= \text{net-}\cancel{\text{flow}}(A, G_k) + \text{net-flow}(A, \bar{A} \setminus G_k)$$

$$\geq \sum_{\substack{i \leq s \\ j \geq l}} \pi_j p_{ji} \, (v_i - v_j)$$

$$\geq (v_s - v_{l+1}) \sum_{\substack{i \leq s \\ j \geq l}} \pi_j p_{ji} \quad \ldots \quad ①$$

By $(***)$

$$\sum_{i=1}^{s} \sum_{j=s+1}^{l} \pi_j p_{ji} \leq \sum_{j=s+1}^{l} \pi_j \leq \frac{\varepsilon \Phi \pi(A)}{4} = \frac{\varepsilon \Phi \gamma_s}{4}$$

Also, by $(**)$ and definition of Φ

$$\sum_{i \in S,\, j \in S^c} \pi_j p_{ji} \geq \Phi \cdot \min(\pi(A), \pi(\bar{A})) \geq \varepsilon \Phi \gamma_s / 2$$

$$\therefore \sum_{\substack{i\in S\\ j\geq l}} \pi_j p_{ji} = \sum_{i\in S^c j} \pi_j p_{ji} - \sum_{i=1}^{s}\sum_{j=s+1}^{l} \pi_j p_{ji}$$

$$\geq \varepsilon \Phi \gamma_s / 4$$

Plug in ①,

$$U_s - U_{l+1} = (u_k - u_{k+1}) \leq \frac{8}{t\varepsilon\Phi\gamma_s}$$

i.e $$(u_k - u_{k+1})\gamma_s \leq \frac{8}{t\varepsilon\Phi}$$

Also $$(u_k - u_{k+1})(\gamma_l - \gamma_s) \leq \frac{8}{t\varepsilon\Phi\gamma_s} \cdot \frac{\varepsilon\Phi\gamma_s}{4} = \frac{2}{t}$$

$\therefore$ lemma is proved.

So now we have found a universal upper bound of the summands, it suffices to find an upper bound of r.

Since $$\sum_{j=k+1}^{l+1} \pi_j > \frac{\varepsilon\Phi\gamma_k}{4}$$

$$\therefore \gamma_{l+1} \geq \left(1+\frac{\varepsilon\Phi}{4}\right)\gamma_k$$

$$\therefore r \leq \log_{\left(1+\frac{\varepsilon\Phi}{4}\right)} (1/\pi_1) + 2$$

$$\leq \frac{4}{\varepsilon\Phi} \ln(1+\pi_1) + 2$$

$$\therefore \frac{1}{t}\left(\frac{8}{\Phi\varepsilon} + 2\right)\left(\frac{4}{\varepsilon\Phi} \ln(1+\pi_1) + 2\right) \leq \frac{\varepsilon}{2}$$

Plug in $t = \dfrac{c \ln(1/\pi_{min})}{\Phi^2 \, \varepsilon^3} \geq \dfrac{c \ln(1/\pi_1)}{\Phi^2 \, \varepsilon^3}$

we arrive to the result.