$
\def*#1{\mathbf{#1}} \def+#1{\mathcal{#1}}
\def-#1{\mathrm{#1}}\def^#1{\mathbb{#1}}\def!#1{\mathtt{#1}}
\newcommand{\norm}[1]{\left\Vert#1\right\Vert}
\newcommand{\abs}[1]{\left\vert#1\right\vert}
\newcommand{\set}[1]{\left{#1\right}}
\newcommand{\tuple}[1]{\left(#1\right)} \newcommand{\eps}{\varepsilon}
\newcommand{\inner}[2]{\langle #1,#2\rangle} \newcommand{\tp}{\tuple}
\renewcommand{\mid}{;\middle\vert;} \newcommand{\cmid}{,:,}
\newcommand{\numP}{#\mathbf{P}} \renewcommand{\P}{\mathbf{P}}
\newcommand{\defeq}{\triangleq} \renewcommand{\d}{,-d}
\newcommand{\ol}{\overline}
\newcommand{\Pr}[2][]{\mathbf{Pr}*{#1}\left[#2\right]}
\newcommand{\E}[2][]{\mathbf{E}*{#1}\left[#2\right]}
\newcommand{\Var}[2][]{\mathbf{Var}_{#1}\left[#2\right]}
\renewcommand{\emptyset}{\varnothing}
$

[Solution of Homework 2]

Problem 1 (Optimal Coupling)

Let Ω be a finite state space and μ,ν be two distributions over Ω. Prove that there exists a coupling ω of μ and ν such that
$
\Pr[(X,Y)\sim \omega]{X\ne Y} = D_{-{TV}}(\mu,\nu).
$
You need to explicitly describe how ω is constructed.
Proof.
Let $P(x,y)$ denote $\Pr[(X,Y)\sim \omega]{X=x,Y=y}$ and use ρ to denote $D_{-{TV}}(\mu,\nu)$ for shorthand. The coupling ω can be constructed as follows: First, we know if we want get the equality, then the diagonal elements must be set as $P(a,a)=\min{\mu(a),\nu(a)},\forall a\in\Omega$. Clearly, if $D_{-{TV}}(\mu,\nu)=0$, the forementioned setting is indeed the optimal coupling. Otherwise, for any $a\in \Omega$, let
\begin{align*
&R_X(a)=\mu(a)-P(a,a)\
&R_Y(a)=\nu(a)-P(a,a).
\end{align*}
It's clear that $\sum_aR_X(a)=\sum_bR_Y(b)=\rho$ by the fact that $D_{-{TV}}(\mu,\nu)=\max_{A\in \Omega}|\mu(A)-\nu(A)|$. Moreover, $R_X(a)R_Y(a)=0, \forall a\in\Omega$.

We should assign the remaining mass to the off-diagonal elements. In fact we only care about the matrix indices in which R_X or R_Y are not 0. If we normalize R_X and R_Y, and we can view R_X/ρ and R_Y/ρ as two distributions. Thus we only need to construct a coupling for these two distributions and the simplest way is just let them be independent. What's more, we should multiply a factor ρ to ensure the summation is still ρ instead of 1. Hence
for any $a,b\in \Omega$ such that $a\neq b$, let $ P(a,b)=\frac{R_X(a)}{\rho}\cdot\frac{R_Y(b)}{\rho}\cdot\rho.$

Obviously P is a legal coupling by above arguments without additional proof. But for some students that are not clear why P looks like this, we can directly verify that P is indeed a legal coupling.

For a fixed $a\in \Omega$,
\begin{align}
|sum_bP(a,b)&=P(a,a)+|sum_{b:b|neq a}P(a,b)|
&=P(a,a)+|sum_{b:b|neq a}|frac{R_X(a)R_Y(b)}{|rho}|
&=P(a,a)+|frac{R_X(a)}{|rho}(|rho-R_Y(a))|
&=P(a,a)+R_X(a)=|mu(a)
|end{align}
Similarly, you can check that $\sum_bP(a,b)=\nu(b),\forall b\in\Omega$. Hence it's a feasible coupling. As for the optimiality,
\begin{align}
|Pr[(X,Y)|sim |omega]{X|ne Y}&=1-|sum_{a|in |Omega} P(a,a)|
&= |sum_{a|in |Omega} |mu(a)- |sum_{a|in |Omega}P(a,a)=|rho.
|end{align}

Problem 2 Total Variation Distance is Non-Increasing

Let P be the transition matrix of an irreducible and aperiodic Markov chain with state space Ω. Let π be its stationary distribution. Let μ_0 be an arbitrary distribution on Ω and $\mu_t^{!T} = \mu_0^{!T}P^t$ for every $t\ge 0$. For every $t\ge 0$, let $\Delta(t) = D_{!{TV}}(\mu_t,\pi)$ be the total variation distance between μ_t and π. Using *coupling* to prove that $\Delta(t+1)\le \Delta(t)$ for every $t\ge 0$.
Proof
Let $X_t\sim \mu_t$ and $Y_t\sim \pi$. By coupling lemma, there exist a coupling ω of X_t and Y_t such that $\Pr{X_t\neq Y_t}=\Delta(t)$. Equipped with ω, we construct the coupling ω' of X_{t+1} and Y_{t+1} as follows:
1. We first sample (X_t,Y_t) from ω;
2. Next we run the Markov chain according to the transition matrix P on X_t and Y_t as follows:

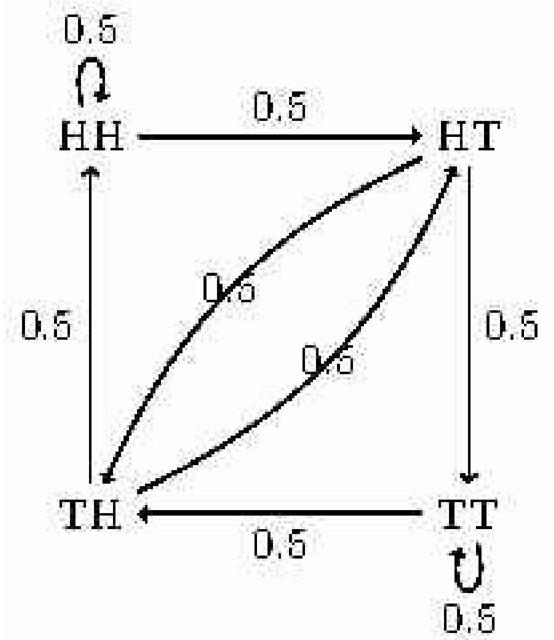
- If $X_t=Y_t$, the two chains evolve synchronously;
- If $X_t\neq Y_t$, the two chains evolve independently.

Hence
\begin{align}
|Delta(t+1)|le |Pr[(X_{t+1},Y_{t+1})|sim |omega^|prime]{X_{t+1}|neq Y_{t+1}}|leq
|Pr[(X_t,Y_t)|sim |omega]{X_{t}|neq Y_{t}}=|Delta(t).
|end{align}

Problem 3

You are tossing a coin repeatedly (In each round, the coin gives H or T with same probability). Which pattern would you expect to see faster: HH or HT? Formally, let T_1 be the first time the pattern HH appears and T_2 be the first time the pattern HT appears. What are $\E{T_1}$ and $\E{T_2}$? (Hint: define a Markov chain with four states HH, HT, TH, TT)

- **bonus**: Can you generalize your results to an arbitrary pattern?
  Solution.



We use $1,2,3,4$ to denote the four states HH, HT, TT, TH for shorthand, respectively. What's more, we use 0 to denote the empty state. For any $0\leq i,j\leq 4$, $T_{i\to j}$ denotes the first time arriving at j when starting from state i. Thus $T_1=T_{0\to 1}$ and $T_2=T_{0\to 2}$.
Since $T_{3\to 4}$ satisfies $!{Geom}(1/2)$, we have $\E{T_{3\to 4}}=2$. Furtheremore, $\E{T_{2\to 4}}=\frac{1}{2}\times 1+\frac{1}{2}\times \tuple{1+\E{T_{3\to 4}}}=2$ and $\E{T_{4\to 1}}=1+\frac{1}{2}\E{T_{2\to 1}}=1+\frac{1}{2}\E{T_{2\to 4}+T_{4\to 1}}=2+\frac{1}{2}\E{T_{4\to 1}}$, which implies $\E{T_{4\to 1}}=4$.
Hence
\begin{align}
\E{T_1}&=\frac{1}{4}\tuple{\E{T_{0\to 1}}+\E{T_{0\to 2}+T_{2\to 1}}+\E{T_{0\to 3}+T_{3\to 1}}+\E{T_{0\to 4}+T_{4\to 1}}+\E{T_{0\to 4}+T_{4\to 1}}}\\
&=2 +\frac{1}{4}\tuple{\E{T_{2\to 4}+T_{4\to 1}}+\E{T_{3\to 4}+T_{4\to 1}}+\E{T_{4\to 1}}}=6.
\end{align}
Symmetrically, we have $\E{T_{4\to 2}}=2$, $\E{T_{1\to 2}}=2$ and $\E{T_{3\to 2}}=\E{T_{3\to 4}+T_{4\to 2}}=4$, So $\E{T_2}=2+\frac{1}{4}\E{T_{1\to 2}+T_{3\to 2}+T_{4\to 2}}=4$.

Therefore, we will expect to see HT faster.

For the bonus, please refer to the slides of Spring 2022 for this course. (http://chihaozhang.com/teaching/SP2022spring/notes/lec7.pdf)

Problem 4 (Path Coupling)

In this problem, we develop the *path coupling* technique to improve our analysis for sampling proper colorings. In the coupling analysis I showed to you in the class, we have to design a coupling for every pair $x,y\in \Omega$. Recall that for two colorings $x,y\in\Omega$, we use $d(x,y)$ to denote their Hamming distance. The path coupling technique allows us to design coupling only for those x,y with $d(x,y)=1$.

Suppose now we already have a coupling satisfying $\E{d(X_{t+1},Y_{t+1})\mid (X_t,Y_t)}\le (1-\alpha) \cdot d(X_t,Y_t)$ for those (X_t,Y_t) with $d(X_t,Y_t)=1$, I will tell you how to extend it to a coupling for arbitrary (X_t,Y_t).

Assuming $d(X,Y)=k$, we can transform the coloring X to the coloring Y by changing their disagreeing vertices one by one. This operation results in $k+1$ intermediate colorings denoted by $X=Z_0,Z_1,Z_2,\dots,Z_k=Y$ where $d(Z_i,Z_{i+1})=1$ for all $i=0,1,\dots,k-1$. Note that some of Z_i might not be proper colorings, and we can slightly extend our state space by adding those improper Z_i. The same transition rule applies to the new states. The new state space might not be irreducible, but it contains an irreducible component containing all proper colorings and the stationary distribution is still the uniform distribution on all proper colorings (verify this!). As before, we assume a coupling satisfying $\E{d(X_{t+1},Y_{t+1})\mid (X_t,Y_t)}\le (1-\alpha) \cdot d(X_t,Y_t)$ for those (X_t,Y_t) with $d(X_t,Y_t)=1$. In particular, we know how to couple (Z_i,Z_{i+1}) for $i=0,1,\dots,k-1$. Use (Z_i',Z_{i+1}') to denote the result of this coupling, namely $(Z_i',Z_{i+1}')=(X_{t+1},Y_{t+1})$ where (X_{t+1},Y_{t+1}) is sampled from the promised coupling when $(X_t,Y_t)=(Z_i,Z_{i+1})$. Consider the following random process:

- Couple (Z_0,Z_1) and obtain $(Z_0',Z_1')=(z_0,z_1)$;
- For $i=1,\dots,k-1$
- Couple (Z_i,Z_{i+1}) to obtain (Z_i',Z_{i+1}') and let z_{i+1} be Z_{i+1}' *conditioned on* $Z_i'=z_i$
- Output (z_0,z_k)

(1) Prove that $(X_{t+1},Y_{t+1})=(z_0,z_k)$ is a legal coupling of (X_t,Y_t).
(2) Prove that in the coupling above, $\E{d(z_0,z_k)}\le (1-\alpha)\cdot k$.
(3) Use the path coupling technique to show that $q>2\Delta$ can guarantee our Markov chain to mix in $O(n\log n)$ steps.

Proof.
The notations in this problem are a little confused, so I will redefine the notations first. Consider we have two chains with states X_t and Y_t. We can transform X_t to Y_t by changing the disagreeing colors one by one to get $X_t=Z_t^0,Z_t^1,Z_t^2,\dots,Z_t^k=Y_t$. By path coupling, we obtain $X_{t+1}=Z_{t+1}^0,Z_{t+1}^1,Z_{t+1}^2,\dots,Z_{t+1}^k=Y_{t+1}$. Let $\omega(X,Y)$ be the promised coupling with the last time state X,Y as the initial condition. Thus $Z_{t+1}^{i+1}=Q$ if $P=Z_{t+1}^i$ when $(P,Q)\sim \omega(Z_t^i,Z_t^{i+1})$.

(1) We only need to prove it for $k=2$, and it is obviously right by induction for the larger k. Our goal is to prove for any coloring z, $\Pr{X_{t+1}=z}=\Pr[(P_1,Q_1)\sim \omega(Z_t^0,Z_t^1)]{P_1=z}$ and $\Pr{Y_{t+1}=z}=\Pr[(P_2,Q_2)\sim \omega(Z_t^1,Z_t^2)]{Q_2=z}$. The first equality is trivial by the definition of the path coupling. Now we prove the second one.
\begin{align*}
\Pr{Y_{t+1}=z}&=\sum_{z^\prime} \Pr{Z_{t+1}^1=z^\prime}\cdot\Pr{Z_{t+1}^2=z | Z_{t+1}^1=z^\prime}\\
&=\sum_{z^\prime}\Pr[(P_2,Q_2)\sim \omega(Z_t^1,Z_t^2)]{P_2=z\prime}\cdot \Pr[(P_2,Q_2)\sim \omega(Z_t^1,Z_t^2)]{Q_2=z|P_2=z^\prime}\\
&=\Pr[(P_2,Q_2)\sim \omega(Z_t^1,Z_t^2)]{Q_2=z}.
\end{align*}
since both X_t and Y_t are running on proper coloring with the metropolis algorithm, thus the stationary distribution of them are both uniform distribution.

In fact, if we reconsider the path coupling techinque with the specific coupling constructed in (3), we know in fact we only toss the dice once for each time t and then randomly choose a vertex and a color to change X_t to X_{t+1}. the remaining $Z_{t+1}^{1},Z_{t+1}^{2},Z_{t+1}^{k}$ are determined by X_t, Y_t and the intermediate state $Z_t^i,1\leq i\leq k$. Obviously it is a legal coupling.

(2) Since $d(Z_t^0,Z_t^k)\leq \sum_{i=0}^{k-1}d(Z_t^i,Z_t^{i+1})$, $\E{d(Z_t^0,Z_t^k)}\leq (1-\alpha)k$.

(3) Now we explicitly define our promised coupling. For any two coloring X_t,Y_t such that $d(X_t,Y_t)=1$, let v be the vertex $X_t(v)\neq Y_t(v)$. We randomly choose a color c and vertex u and try to recolor $X_t(u)$ with c. Consider the case that u is a neighbor of v. If $q=X_t(v)$, then recolor $Y_t(u)$ with $Y_t(v)$; if $q=Y_t(v)$, then recolor $Y_t(u)$ with $X_t(v)$. For the other case, just try to recolor $Y_t(u)$ with c.
Now similar to the analysis of the lecture note, the good move happens when $u=v$ and c recolors $X_t$and Y_t successfully:$\Pr{d(X_{t+1},Y_{t+1})=d(X_{t+1},Y_{t+1})-1}\geq \frac{1}{n}\cdot\frac{q-\Delta}{q};$
the bad move means u is a neighbor of v and $c=Y_t(v)$:
$\Pr{d(X_{t+1},Y_{t+1})=d(X_{t+1},Y_{t+1})+1}\leq \frac{\Delta}{n}\cdot\frac{1}{q}.$

Therefore,
\begin{aligned}
&\mathbf{E}\left[d\left(X_{t+1}, Y_{t+1}\right) \mid\left(X_t, Y_t\right)\right] \
& =d\left(X_t, Y_t\right)+\operatorname{Pr}\left[d\left(X_{t+1}, Y_{t+1}\right)=d\left(X_t, Y_t\right)+1\right]-\operatorname{Pr}\left[d\left(X_{t+1}, Y_{t+1}\right)=d\left(X_t, Y_t\right)-1\right] \
& \leq d\left(X_t, Y_t\right)+\frac{\Delta d\left(X_t, Y_t\right)}{n} \cdot \frac{1}{q}-\frac{d\left(X_t, Y_t\right)}{n} \cdot \frac{q- \Delta}{q} \
& \leq d\left(X_t, Y_t\right)\left(1-\frac{q-2 \Delta}{n q}\right)
\end{aligned}
In the case $q>2\Delta$, if we want $D_{\mathrm{TV}} \leq\left(1-\frac{1}{n q}\right)^t n \leq \varepsilon$, we have the mixing time is bounded by $\tau_{\mathrm{mix}}(\varepsilon) \leq n q \log \frac{n}{\varepsilon}$.

