Advanced Algorithms (XIII)

Shanghai Jiao Tong University

Chihao Zhang

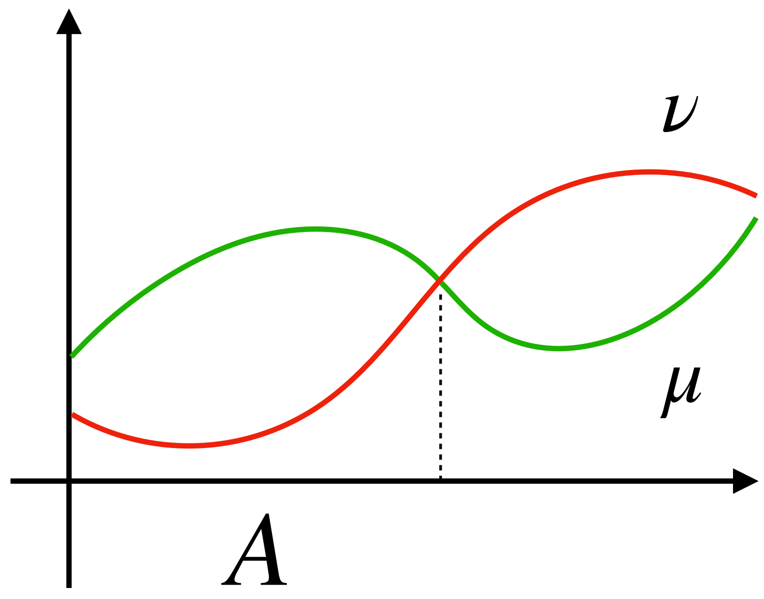
June 1, 2020

Total Variation Distance

Let μ and ν be two distributions on Ω

Their total variation distance is

$$d_{\mathrm{TV}}(\mu, \nu) = \frac{1}{2} \sum_{x \in \Omega} \left| \mu(x) - \nu(x) \right| = \max_{A \subseteq \Omega} \mu(A) - \nu(A)$$



ℓ_1-distance scaled by $\frac{1}{2}$

Coupling

Let μ and ν be two distributions on Ω

A coupling of μ and ν is a joint distribution ω on $\Omega \times \Omega$ such that:

$$\forall x \in \Omega, \quad \mu(x) = \sum_{y \in \Omega} \omega(x, y)$$

$$\forall y \in \Omega, \quad \nu(x) = \sum_{x \in \Omega} \omega(x, y)$$

Coupling Lemma

Let ω be a coupling of μ and ν

$(X, Y) \sim \omega \implies X \sim \mu$ and $Y \sim \nu$

Then $$\Pr_{(X,Y)\sim\omega}[X \neq Y] \geq d_{\mathrm{TV}}(\mu, \nu)$$

Moreover, there exists ω^* such that

$$\Pr_{(X,Y)\sim\omega^*}[X \neq Y] = d_{TV}(\mu, \nu)$$

Proof of Coupling Lemma

For finite Ω, designing a coupling is equivalent to filling a $\Omega \times \Omega$ matrix so that the marginals are correct

$$\Omega = \{1,2\},\ \mu = (1/2,1/2),\ \nu = (1/3,2/3)$$

$\nu \backslash \mu$	$\frac{1}{2}$	$\frac{1}{2}$
$\frac{1}{3}$	$\frac{1}{3}$	0
$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{2}$

ω^* is the one maximizing the sum of diagonals

Coupling of Markov Chains

Consider two copies of the chain P:

- The initial distribution is μ_0 and ν_0
- $\mu_t^T = \mu_0^T P^t$ and $\nu_t^T = \nu_0^T P^t$

A coupling of the two chains is joint distribution ω of $\{\mu_t\}_{t\geq 0}$ and $\{\nu_t\}_{t\geq 0}$ satisfying the following conditions

$\{(X_t, Y_t)\}_{t \geq 0} \sim \omega$ is a pair of processes such that

$$\forall a, b \in \Omega, \mathbf{Pr}[X_{t+1} = b \mid X_t = a] = P(a, b)$$

$$\forall a, b \in \Omega, \mathbf{Pr}[Y_{t+1} = b \mid X_t = a] = P(a, b)$$

Marginally, $\{X_t\}$ and $\{Y_t\}$ are both chain P

$$\forall t \geq 0, X_t = Y_t \implies X_{t'} = Y_{t'} \text{ for all } t' > t$$

Two chains coalesce once they meet

Fundamental Theorem via Coupling

If a finite chain P is irreducible and aperiodic, then it has a unique stationary distribution π. Moreover, for any initial distribution μ, it holds that

$$\lim_{t\to\infty} \mu^T P^t = \pi^T$$

Consider two chains $\{X_t\}_{t\geq 0}$ and $\{Y_t\}_{t\geq 0}$

- $X_0 \sim \pi$, $Y_0 \sim \mu_0$ for arbitrary μ_0
- A coupling where X_t and Y_t run independently

$$\text{irreducible} + \text{aperiodic} \implies \exists t, \forall x, y, P^t(x, y) > 0$$

Then for any $z \in \Omega$, there exists some $\theta > 0$ s.t.

$$\begin{aligned}\mathbf{Pr}[X_t = Y_t] \geq \mathbf{Pr}[X_t = Y_t = z] &= \mathbf{Pr}[X_t = z] \cdot \mathbf{Pr}[Y_t = z] \\ &= \pi(z) \cdot P^t(Y_0, z) \geq \theta > 0\end{aligned}$$

$$\mathbf{Pr}[X_t \neq Y_t] \leq 1 - \theta < 1$$

$$\begin{aligned}\mathbf{Pr}[X_{2t} \neq Y_{2t}] &= \mathbf{Pr}[X_{2t} \neq Y_{2t} \wedge X_t = Y_t] + \mathbf{Pr}[X_{2t} \neq Y_{2t} \wedge X_t \neq Y_t] \\ &= \mathbf{Pr}[X_{2t} \neq Y_{2t} \mid X_t \neq Y_t] \cdot \mathbf{Pr}[X_t \neq Y_t] \\ &\leq (1 - \theta)^2\end{aligned}$$

$$\cdots$$

$$\mathbf{Pr}[X_{kt} \neq Y_{kt}] \leq (1 - \theta)^k$$

$$\lim_{n \to \infty} \mathbf{Pr}[X_n \neq Y_n] = 0$$

Mixing Time

The mixing time $\tau_{\text{mix}}(\varepsilon)$ is the the first time t such that the total variation distance between X_t and π is at most ε, for any initial X_0

$$\tau_{\text{mix}}(\varepsilon) = \max_{\mu_0} \min_{t \geq 0} d_{\text{TV}}(\mu_0^T P^t, \pi) \leq \varepsilon$$

$$\tau_{\text{mix}} = \tau_{\text{mix}}(1/4)$$

Random Walk on Hyper Cube

- $V = \{0,1\}^n$
- $x \sim y$ iff $\|x - y\|_1 = 1$

Lazy walk on G

Standing at $x \in \{0,1\}^n$

- with prob. $\frac{1}{2}$, do nothing
- otherwise, choose $i \in [n]$ u.a.r and flip $x(i)$

The chain is equivalent to

- choose $i \in [n]$ and $b \in \{0,1\}$ u.a.r.
- change $x(i) \leftarrow b$

Let X_t and Y_t be two walks

We couple them by choosing the same i and b

What is the probability that $X_t \neq Y_t$?

Coupon Collector!

If $t \geq n \log n + cn$, then $\mathbf{Pr}[X_t \neq Y_t] \leq e^{-c}$

Coupling lemma implies that

$$\tau_{\text{mix}}(\varepsilon) \leq n \log n + n \log \varepsilon^{-1}$$

Another Random Walk

Lazy walk on G

Standing at $x \in \{0,1\}^n$

- with prob. $\dfrac{1}{n+1}$, do nothing
- otherwise, choose $i \in [n]$ u.a.r and flip $x(i)$

A coupling argument implies $\tau_{\mathrm{mix}} \leq \dfrac{1}{2} n \log n + O(n)$

Reversible Chain

Recall that we say a Markov chain P is reversible with respect to π if

$$\forall x, y \in \Omega, \quad \pi(x)P(x,y) = \pi(y)P(y,x)$$

Then π is a stationary distribution of P

We showed that spectral decomposition is a powerful tool to analyze reversible chains

Relaxation Time

$$P = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T D_\pi$$

$$P^t = \sum_{i=1}^{n} \lambda_i^t \mathbf{v}_i \mathbf{v}_i^T D_\pi$$

- $\lambda_n = 1$
- $\lambda_1 \geq -1$ and $\lambda_1 = -1$ if and only if P is bipartite
- $\lambda_{n-1} = 1$ if and only P is reducible

If P is reducible ($\lambda_{n-1} < 1$) and aperiodic ($\lambda_1 > -1$)

$$\lim_{t\to\infty} P^t = \mathbf{1}\mathbf{1}^T D_\pi = \begin{bmatrix} \pi^T \\ \pi^T \\ \vdots \\ \pi^T \end{bmatrix}$$

$$\lambda_\star := \max_{1 \leq i \leq n-1} |\lambda_i|$$

$$\tau_{\mathrm{rel}} := \frac{1}{1 - \lambda_\star}$$

For reversible, irreducible, aperiodic chains:

$$\pi_{\min} := \min_x \pi(x)$$

$$(\tau_{\mathrm{rel}} - 1)\log\left(\frac{1}{2\varepsilon}\right) \leq \tau_{\mathrm{mix}}(\varepsilon) \leq \tau_{\mathrm{rel}} \log\left(\frac{1}{\varepsilon \pi_{\min}}\right)$$