Spectral gap for nonreversible Markov chains

Sourav Chatterjee

Reversible Markov chains

- Let $X_0, X_1, \ldots$ be a time-homogeneous stationary Markov chain on a finite state space $\mathcal{S}$.
- Let P be the transition matrix of the chain.
- Let μ be an invariant probability measure of our chain, such that $\mu(x) > 0$ for all $x \in \mathcal{S}$.
- The measure μ defines an inner product on $\mathbb{C}^{\mathcal{S}}$ as

$$\langle f, g \rangle := \sum_{x \in \mathcal{S}} f(x) \overline{g(x)} \mu(x).$$

- The Markov chain is said to be reversible if P is self-adjoint with respect to this inner product.

Spectral gap of a reversible Markov chain

- If the chain is reversible, then P has eigenvalues $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{|\mathcal{S}|} \geq -1$ (repeated by multiplicities).
- The spectral gap is defined to be $1 - \lambda_2$.
- The absolute spectral gap is $1 - \max\{\lambda_2, |\lambda_{|\mathcal{S}|}|\}$.
- If $P(x,x) \geq 1/2$ for all x, the chain is called strongly aperiodic.
- Any chain can be made strongly aperiodic by decreeing that at each step, it stays where it is with probability $1/2$.
- For strongly aperiodic chains, $\lambda_{|\mathcal{S}|} \geq 0$, and hence the spectral gap is equal to the absolute spectral gap.
- The relaxation time τ_{rel} for a reversible chain is defined to be the reciprocal of the absolute spectral gap.

A characterization of the relaxation time

- For reversible chains, the relaxation time can be characterized as the time required for decorrelation of L^2 functions.
- To be precise, if $X_0, X_1, \ldots$ is a stationary reversible Markov chain with invariant distribution μ and relaxation time τ_{rel}, then for any real-valued $f, g \in L^2(\mu)$ and any n,

  $$\mathrm{Corr}(f(X_0), g(X_n)) \le \left(1 - \frac{1}{\tau_{\mathrm{rel}}}\right)^n,$$

  and equality is attained if we take $f = g =$ the eigenvector of P corresponding to the eigenvalue with the second largest magnitude (i.e., either λ_2 or $\lambda_{|\mathcal{S}|}$).
- The spectral gap and the relaxation time have many other important properties and are central to the analysis of reversible Markov chains.

Spectral gap of nonreversible chains?

- There have been many attempts at defining spectral gap for nonreversible chains, but none that is widely accepted like the one for reversible chains.
- For example, one popular definition of spectral gap for nonreversible chains is one minus the second largest singular value of P.
- If γ denotes this spectral gap, then the chain has the contraction property that for any $f \in L^2(\mu)$ with $\mu f := \sum f(x)\mu(x) = 0$,

$$\|Pf\| \le (1-\gamma)\|f\|,$$

  where $\|\cdot\|$ is the $L^2(\mu)$ norm.
- However, if one defines the relaxation time to be the inverse of this spectral gap, it may lead to misleading pictures in various situations.

Example

- The Chung–Diaconis–Graham chain is a Markov chain on $\mathbb{Z}/N\mathbb{Z}$ which proceeds as

$$X_{i+1} = 2X_i + \varepsilon_{i+1} \bmod N,$$

where $\varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. random variables that are uniformly distributed in $\{-1, 0, 1\}$.
- The second largest singular value of P is like $1 - c/N^2$, and therefore the relaxation time, according to the definition in the previous slide, is of order N^2.
- However, the mixing time (definition later) of this chain is at most of order $\log N \log\log N$. That is, from any starting state, X_i is close to being uniformly distributed on $\mathbb{Z}/N\mathbb{Z}$ whenever $i \gg \log N \log\log N$.

A new definition of spectral gap for nonreversible chains

- In the recently posted preprint titled "Spectral gap of nonreversible Markov chains", I have proposed the following definition of spectral gap for nonreversible chains.
- Given a Markov chain on a finite state space with transition matrix P, I define the spectral gap γ to be the second smallest singular value of the generator $L := I - P$.
- Note that unless the chain is reversible, there is no obvious relation between this definition of spectral gap and the one mentioned earlier.
- I define the relaxation time to be the reciprocal of this spectral gap. Henceforth, this relaxation time will be denoted by τ.

Main result, informally

- Choose the starting state $X_0 \sim \mu$, and define

$$\Delta_n := \sup_{g\,:\,\|g-\mu g\|=1} \left\{ \mathbb{E}\left[\left(\frac{1}{n} \sum_{i=0}^{n-1} g(X_i) - \mu g \right)^2 \right] \right\}^{1/2}.$$

- The main theorem of the preprint, displayed in the next slide, shows that Δ_n is small if and only if $n \gg \tau$.
- This gives a characterization of our relaxation time as the time required for convergence of time averages to their limiting values in L^2.

Main result, formal statement

Theorem (C., 2023)

For any $n \geq 1$,

$$\Delta_n \leq \sqrt{\frac{4\tau}{n}}.$$

Conversely, for any $n \leq \tau/3$,

$$\Delta_n \geq \frac{1}{132}.$$

Contraction is not necessary

- Consider the Markov chain on $\mathbb{Z}/N\mathbb{Z}$ that always moves one step to the right, i.e.,

$$X_{i+1} = X_i + 1 \bmod N.$$

- This chain never mixes (and in particular, P is not contracting and correlations do not tend to zero), but the relaxation time (according to our definition) is of order N.
- A moment's thought reveals that this indeed makes sense, since empirical averages converge in time $\gg N$.

The Chung–Diaconis–Graham chain

- In the preprint, it is shown that our relaxation time for the Chung–Diaconis–Graham chain is of order $\log N$ up to constants, whenever N is a prime.
- Thus, empirical averages converge at time $\log N$ whenever N is a prime.
- The mixing time of the chain, for a Mersenne prime N (i.e., a prime of the form $2^n - 1$), is known to be exactly of order $\log N \log \log N$, and for other primes, to be at most of this order.
- More examples later.

Proof of the main result: Upper bound

- Take any $g : \mathcal{S} \to \mathbb{R}$ such that $\mu g = 0$ and $\|g\| = 1$.
- For any $i \le j$,

$$\begin{aligned}\mathbb{E}(g(X_i)g(X_j)) &= \mathbb{E}(g(X_i)\mathbb{E}(g(X_j)|X_i)) \\ &= \mathbb{E}(g(X_i)P^{j-i}g(X_i)) = \langle g, P^{j-i}g\rangle.\end{aligned}$$

- Thus,

$$\begin{aligned}\mathbb{E}\biggl[\biggl(\frac{1}{n}\sum_{i=0}^{n-1} g(X_i)\biggr)^2\biggr] &\le \frac{2}{n^2}\sum_{0\le i\le j\le n-1}\langle g, P^{j-i}g\rangle \\ &= \frac{2}{n^2}\sum_{i=0}^{n-1}\langle g, h_{n-i}\rangle,\end{aligned}$$

where

$$h_k := \sum_{m=0}^{k-1} P^m g.$$

Upper bound, contd.

- By Cauchy–Schwarz,

$$\langle g, h_k \rangle \le \|g\| \|h_k\| = \|h_k\|.$$

- Thus, we arrive at

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=0}^{n-1} g(X_i)\right)^2\right] \le \frac{2}{n^2}\sum_{k=0}^{n-1} \|h_k\|,$$

where

$$h_k = \sum_{m=0}^{k-1} P^m g.$$

Upper bound, contd.

- Let f be the solution of
$$Lf = g.$$
(Recall that $L = I - P$, and τ is the inverse of the second smallest singular value of L.)
- Since $\mu g = 0$ and $\|g\| = 1$, we get
$$1 = \|g\| = \|Lf\| \geq \frac{1}{\tau}\|f\|.$$
- But, note that
$$h_k = \sum_{m=0}^{k-1} P^m g = \sum_{m=0}^{k-1} P^m (I - P) f = f - P^k f.$$
- Thus, for any k,
$$\|h_k\| \leq \|f\| + \|P^k f\| \leq 2\|f\| \leq 2\tau.$$
- This is the main inequality. The above trick allows us to replace contraction by cancellation.

Upper bound, contd.

- This gives us

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=0}^{n-1} g(X_i)\right)^2\right] \le \frac{2}{n^2}\sum_{k=0}^{n-1}\|h_k\| \le \frac{2}{n^2}\sum_{k=0}^{n-1} 2\tau = \frac{4\tau}{n},$$

completing the proof of the upper bound.

Lower bound

- Take any f that minimizes $\|Lf\|$ subject to $\mu f = 0$ and $\|f\| = 1$.
- Let $g := Lf$. Note that $\|g\| = \gamma$ and $\mu g = 0$.
- The main idea is that if n is smaller than our relaxation time, then *either the time averages of g or those of f will not converge by time n.*
- For each n, define

$$u_n := f - \frac{1}{n}\sum_{k=n}^{2n-1} P^k f, \quad v_n := \frac{1}{n}\sum_{k=n}^{2n-1} P^k f.$$

- Then for any n,

$$1 = \|f\| = \|u_n + v_n\| \leq \|u_n\| + \|v_n\|.$$

- Recall:

$$\Delta_n = \sup_{h\,:\,\mu h=0,\, \|h\|=1} \left\{\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=0}^{n-1} h(X_i)\right)^2\right]\right\}^{1/2}.$$

Lower bound, contd.

- Now,

$$u_n = \frac{1}{n}\sum_{k=n}^{2n-1}(f - P^k f) = \frac{1}{n}\sum_{k=n}^{2n-1}\sum_{m=0}^{k-1} P^m (I-P)f$$
$$= \frac{1}{n}\sum_{k=n}^{2n-1}\sum_{m=0}^{k-1} P^m g = \frac{1}{n}\sum_{k=n}^{2n-1} h_k.$$

- But since $P^m g(X_0) = \mathbb{E}(g(X_m)|X_0)$,

$$\|h_k\| = \biggl\{\mathbb{E}\biggl[\biggl(\sum_{m=0}^{k-1} P^m g(X_0)\biggr)^2\biggr]\biggr\}^{1/2}$$
$$\le \biggl\{\mathbb{E}\biggl[\biggl(\sum_{m=0}^{k-1} g(X_m)\biggr)^2\biggr]\biggr\}^{1/2} \le k\Delta_k \|g\| = k\gamma\Delta_k.$$

- Thus, $\|u_n\| \le 2n\gamma \max_{n\le k\le 2n-1} \Delta_k$.
- Intuitively, $\|u_n\|$ is small if the time average of g is small at time n.

Lower bound, contd.

- On the other hand,

$$\begin{aligned}\|v_n\| &= \left\| \frac{1}{n} \sum_{k=n}^{2n-1} P^k f \right\| \\ &= \left\| \frac{1}{n} \sum_{k=0}^{2n-1} P^k f - \frac{1}{n} \sum_{k=0}^{n-1} P^k f \right\| \\ &\le \left\| \frac{1}{n} \sum_{k=0}^{2n-1} P^k f \right\| + \left\| \frac{1}{n} \sum_{k=0}^{n-1} P^k f \right\| \\ &\le 2\Delta_{2n}\|f\| + \Delta_n\|f\| \le 3\max\{\Delta_n, \Delta_{2n}\},\end{aligned}$$

  since $\|f\| = 1$.
- Again, intuitively, $\|v_n\|$ is small if the time average of f is small at time n.

Lower bound, contd.

- Combining, we get that for any n,

$$\begin{aligned} 1 &\leq \|u_n\| + \|v_n\| \\ &\leq 2n\gamma \max_{n \leq k \leq 2n-1} \Delta_k + 3\max\{\Delta_n, \Delta_{2n}\} \\ &\leq (2n\gamma + 3) \max_{n \leq k \leq 2n} \Delta_k. \end{aligned}$$

- In particular, if $n \leq 4\tau = 4/\gamma$, then there is some k between n and $2n$ such that $\Delta_k \geq 1/11$.
- It is not hard to show that Δ_n satisfies the subadditive inequality

$$\Delta_{n+m} \leq \frac{n}{n+m}\Delta_n + \frac{m}{n+m}\Delta_m.$$

- Combining this with the previous observation, it is not hard to show that if $n \leq \tau/3$, then $\Delta_n \geq 1/132$ (I will spare you the details).

Example: Local random walks on tori

- Take any $d \geq 1$. Let $(p_i)_{-d \leq i \leq d}$ be a set of nonnegative real numbers that sum to 1.
- Consider the random walk on $(\mathbb{Z}/N\mathbb{Z})^d$ which takes independent steps as $X_n = X_{n-1}$ with probability p_0 and $X_n = X_{n-1} \pm e_i$ with probability $p_{\pm i}$ for $i = 1, \ldots, d$, where $e_1, \ldots, e_d$ are the standard basis vectors of $\mathbb{R}^d$.
- We assume that $p_i + p_{-i} > 0$ for each $1 \leq i \leq d$, because otherwise the Markov chain is not irreducible.
- Under this condition, the uniform distribution μ on the torus is the unique invariant measure of the walk.
- It is not hard to show that the mixing time of the chain is of order N^2 for any choice of p_i's satisfying the above constraints.

Example, contd.

Theorem (C., 2023)

The relaxation time τ (according to our definition) of the above random walk is at most of order N^2, for any choice of p_i's and any d. Moreover:

- *If $d = 1$, then τ is of order N^2 if the chain is reversible, and of order N if it is not.*
- *If $d \geq 2$, and $p_i = p_{-i}$ for some i or $p_i - p_{-i}$ is rational for at least two i's, then τ is at least of order N^2.*

Can the upper bound on τ be made better than N^2 with some choice of p_i's if $d \geq 2$?

Example, contd.

Theorem (C., 2023)

For any d and any choice of p_i's, τ is at least of order $N^{2d/(d+1)}$. Moreover, for almost every choice of p_i's (with respect to Lebesgue measure on the simplex), we have $\tau = N^{2d/(d+1)+o(1)}$ as $N \to \infty$.

What about an explicit choice that attains the lower bound?

Theorem (C., 2023)

Let $\alpha \in (0,1)$ be an irrational algebraic number of degree two, such as $1/\sqrt{2}$. Consider the random walk on $(\mathbb{Z}/N\mathbb{Z})^2$ which moves either one step up with probability $1-\alpha$, or one step to the right with probability α. The relaxation time of this walk (according to our definition) satisfies $C_1 N^{4/3} \le \tau \le C_2 N^{4/3}$, where C_1 and C_2 are positive constants that depend only on α.

I will sketch the proof of this theorem if I have time at the end.

Cheeger constant

- Let $X_0, X_1, \ldots$ be a stationary Markov chain on a finite state space $\mathcal{S}$ with transition matrix P and invariant measure μ.
- The Cheeger constant of the chain is defined as

$$\xi := \min_{A \subseteq \mathcal{S},\, 0 < \mu(A) \le 1/2} \mathbb{P}(X_1 \notin A | X_0 \in A).$$

- A famous property of the Cheeger constant is that if the Markov chain is reversible with spectral gap γ, then

$$\frac{\xi^2}{2} \le \gamma \le 2\xi,$$

and equalities are attained on both sides (up to constants).

Generalization to nonreversible chains

Theorem (C., 2023)

Let $X_0, X_1, \ldots$ be a Markov chain on a finite state space with invariant measure μ and spectral gap γ, according to our definition. Let ξ be the Cheeger constant of the chain. Then

$$\frac{\xi^2}{16} \le \gamma \le 32\xi.$$

- No other definition of spectral gap for nonreversible chains satisfies this kind of relation with the Cheeger constant.
- The proof is quite different than the reversible case, since one can no longer work with eigenvectors.

Mixing time

- The total variation distance between two probability measures μ and ν on the finite set $\mathcal{S}$ is defined as

$$\|\mu - \nu\|_{\mathrm{TV}} := \max_{A \subseteq \mathcal{S}} |\mu(A) - \nu(A)|.$$

- Given $\varepsilon > 0$, the mixing time $\tau_{\mathrm{mix}}(\varepsilon)$ of a Markov chain with stationary distribution μ is defined as

$$\tau_{\mathrm{mix}}(\varepsilon) := \inf\{n \geq 0 : \|P_x^n - \mu\|_{\mathrm{TV}} \leq \varepsilon\},$$

  where P_x^n is the law of X_n given $X_0 = x$.
- For reversible chains, the relaxation time and the mixing time satisfy the pair of inequalities

$$(\tau_{\mathrm{rel}} - 1) \log \frac{1}{2\varepsilon} \leq \tau_{\mathrm{mix}}(\varepsilon) \leq \tau_{\mathrm{rel}} \log \frac{1}{\varepsilon \mu_{\min}},$$

  where $\mu_{\min} := \min_{x \in \mathcal{S}} \mu(x)$ and $\varepsilon \in (0, 1/2)$.

Generalization to nonreversible chains

Theorem (C., 2023)

Take any $\varepsilon \in (0, 1/5)$. *Let* τ *be the relaxation time according to our definition. Then*

$$\tau \le \left(\frac{4}{\log(2/(1+4\varepsilon+2\varepsilon^2))} + 2\right)\tau_{\mathrm{mix}}(\varepsilon).$$

Moreover, if $P(x,x) \ge 1/2$ *for all* x, *then for any* $\varepsilon \in (0, 1/2)$,

$$\tau_{\mathrm{mix}}(\varepsilon) \le 1 + 12\tau^2 \log \frac{1}{2\varepsilon\mu_{\mathrm{min}}}.$$

Remarks

- For the second inequality, the condition $P(x, x) \geq 1/2$ cannot be dropped. For example, the chain on $\mathbb{Z}/N\mathbb{Z}$ that always moves one step to the right has $\tau_{\text{mix}}(\varepsilon) = \infty$ for all $\varepsilon \in (0, 1/2)$, but finite τ.
- Note that the upper bound has τ^2 instead of τ as we had in the reversible case. This, too, cannot be improved.
- For example, consider the chain on $\mathbb{Z}/N\mathbb{Z}$ which evolves as

$$X_{i+1} = \begin{cases} X_i + 1 \bmod N & \text{with probability } 1/2, \\ X_i & \text{with probability } 1/2. \end{cases}$$

- This chain mixes in time N^2, but τ is of order N.

Relation to spectral gaps of reversibilized chains

- Let P be the transition matrix of a Markov chain on the finite state space with invariant measure μ.
- Let P^* denote the adjoint of P with respect to the inner product induced by μ.
- The multiplicative reversibilization of is this chain is the reversible Markov chain with transition matrix $M = PP^*$, and the additive reversibilization is the chain with transition matrix $A = \frac{1}{2}(P + P^*)$.
- It is not hard to show that P^*, A and M are Markov transition matrices, and that μ is an invariant measure for all three.

Theorem (C., 2023)

Let γ be the spectral gap of the Markov chain with transition matrix P (according to our definition), and let γ_M and γ_A be the spectral gaps of the multiplicative and additive reversibilizations of P, defined above. Then $\frac{1}{2}\gamma_A \le \gamma \le \sqrt{6\gamma_A}$ and $\frac{1}{2}\gamma_M \le \gamma$. Moreover, if $P(x,x) \ge 1/2$ for $x \in \mathcal{S}$, then we also have $\gamma \le \sqrt{6\gamma_M}$. The inequalities are sharp up to constants.

Summary

- I gave a definition of spectral gap for general Markov chains on finite state spaces that reduces to the standard definition in the reversible case.
- The main result is that the relaxation time, defined to be the inverse of this spectral gap, can be characterized as the time required for empirical averages to converge. Arguably, this notion of convergence is more relevant for practical applications of Markov chains than the mixing time.
- Auxiliary results show that this spectral gap is related to the Cheeger constant and the mixing time via inequalities that are similar to the reversible case.
- Examples show that this relaxation time can sometimes be substantially smaller than the mixing time of the chain if the chain is nonreversible.
- For details of the proofs, and further results, examples, and references, see my preprint on arXiv.

Proof of theorem about random walk on $(\mathbb{Z}/N\mathbb{Z})^2$

Recall:

Theorem (C., 2023)

Let $\alpha \in (0,1)$ be an irrational algebraic number of degree two, such as $1/\sqrt{2}$. Consider the random walk on $(\mathbb{Z}/N\mathbb{Z})^2$ which moves either one step up with probability $1-\alpha$, or one step to the right with probability α. The relaxation time of this walk (according to our definition) satisfies $C_1 N^{4/3} \le \tau \le C_2 N^{4/3}$, where C_1 and C_2 are positive constants that depend only on α.

Also recall: The relaxation time τ is the inverse of the spectral gap γ, which is defined to be the second smallest singular value of $L = I - P$.

Proof of the lower bound

- Let M be a positive integer, to be chosen later.
- Consider the numbers $a\alpha + b(1-\alpha)$ where a, b range over all integers in $[-M, M]$.
- These numbers are all in $[-M, M]$.
- There are $(2M+1)^2$ choices of (a, b), and $[-M, M]$ can be partitioned into $(2M)^2$ sub-intervals of length $(2M)^{-1}$.
- Therefore, by the pigeon hole principle, there exist distinct (a, b) and (c, d) such that $a\alpha + b(1-\alpha)$ and $c\alpha + d(1-\alpha)$ belong to the same sub-interval.
- Let $k := a - c$ and $m := b - d$.

Proof of the lower bound, contd.

- Thus, k and m are integers in $[-2M, 2M]$, not both zero, such that

$$|k\alpha + m(1-\alpha)| \le \frac{1}{2M}.$$

- For $(x, y) \in (\mathbb{Z}/N\mathbb{Z})^2$, let

$$f(x, y) := e^{2\pi \mathrm{i}(kx+my)/N}.$$

- It is easy to check that $\mu f = 0$, $\|f\| = 1$ and

$$\begin{aligned} Pf(x) &= (\alpha e^{2\pi \mathrm{i} k/N} + (1-\alpha)e^{2\pi \mathrm{i} m/N})f(x) \\ &= f(x) + \frac{2\pi \mathrm{i}}{N}(k\alpha + m(1-\alpha))f(x) + O(M^2N^{-2})f(x) \\ &= (1 + O(M^{-1}N^{-1}) + O(M^2N^{-2}))f(x). \end{aligned}$$

- Taking $M = N^{1/3}$ gives $\|(I-P)f\| = O(N^{-4/3})$, showing that τ is at least of order $N^{4/3}$.

Proof of the upper bound

- By an explicit calculation, it can be shown that the singular values of the generator of this chain are

$$\gamma_{k,m} := |1 - \alpha e^{2\pi i k/N} - (1-\alpha) e^{2\pi i m/N}|,$$

where k, m range in $\{0, 1, \ldots, N-1\}$.

- Using this formula and a second order Taylor approximation, it can be shown that to prove the upper bound $\tau = O(N^{4/3})$, it suffices to show that

$$\min_{\substack{0 \le k, m \le N-1, \\ k \ne 0 \text{ or } m \ne 0}} \left(\left| \frac{1}{N}(\alpha k + (1-\alpha)m) \right| + \frac{1}{N^2}(\alpha k^2 + (1-\alpha)m^2) \right) \ge C N^{-4/3},$$

where C is a positive constant that does not depend on N.

Proof of the upper bound, contd.

- Since α solves a quadratic equation with integer coefficients, we can write
$$\alpha = \frac{a + s\sqrt{b}}{c}$$
where $a, b, c \in \mathbb{Z}$, $s \in \{-1, 1\}$, $b \geq 0$, and $c \neq 0$.
- Since α is irrational, b is not a perfect square.
- Take any two integers p, q, at least one of which is nonzero.
- Since b is not a perfect square and p, q are integers which are not both zero, we must have $|q^2 b - p^2| \geq 1$.
- But
$$|q^2 b - p^2| = |q\sqrt{b} - p||q\sqrt{b} + p| \leq C|q\sqrt{b} - p|(|p| + |q|),$$
where C denotes a constant that depends only on α.
- Thus,
$$|q\sqrt{b} - p| \geq \frac{1}{C(|p| + |q|)}.$$

Proof of the upper bound, contd.

- Now, for any $k, m \in \mathbb{Z}$,

$$|\alpha k + (1-\alpha)m| = \left|\frac{a + s\sqrt{b}}{c}(k-m) + m\right| = \frac{|s(k-m)\sqrt{b} + a(k-m) + mc|}{|c|}.$$

- If at least one of k, m is nonzero, then at least one of $s(k-m)$ and $a(k-m) + mc$ is nonzero.
- Thus, we get

$$|\alpha k + (1-\alpha)m| \geq \frac{C_1}{|k-m| + |m|} \geq \frac{C_2}{|k| + |m|}.$$

Proof of the upper bound, contd.

- This gives

$$
\begin{aligned}
&\left|\frac{1}{N}(\alpha k + (1-\alpha)m)\right| + \frac{1}{N^2}(\alpha k^2 + (1-\alpha)m^2) \\
&\ge \frac{C_1}{N(|k|+|m|)} + \frac{C_2(|k|+|m|)^2}{N^2} \\
&\ge C_3\left(\frac{1}{N(|k|+|m|)}\right)^{2/3}\left(\frac{(|k|+|m|)^2}{N^2}\right)^{1/3} = \frac{C_3}{N^{4/3}},
\end{aligned}
$$

where the second step follows by Young's inequality.