# Some Final Problems of
# Alibaba Global Mathematics Competition

## Algebra & Number Theory

### Question 1

Let $p$ be an odd prime number, and let $m \geq 0$ and $N \geq 1$ be integers. Let $\Lambda$ be a free $\mathbb{Z}/p^N\mathbb{Z}$-module of rank $2m+1$, equipped with a perfect symmetric $\mathbb{Z}/p^N\mathbb{Z}$-bilinear form

$$(\ ,\ ):\ \Lambda \times \Lambda \to \mathbb{Z}/p^N\mathbb{Z}.$$

Here "perfect" means that the induced map

$$\Lambda \to \mathrm{Hom}_{\mathbb{Z}/p^N\mathbb{Z}}(\Lambda, \mathbb{Z}/p^N\mathbb{Z}), \quad x \mapsto (x, \cdot)$$

is an isomorphism. Find the cardinality of the set

$$\big\{x \in \Lambda \mid (x,x) = 0\big\},$$

expressed in term of $p, m, N$.

### Answer 1

For every integer $0 \leq n \leq N$, we put

$$\Lambda(n) := \{x \in \Lambda \mid (x,x) \in p^n\mathbb{Z}/p^N\mathbb{Z}\},$$

which is stable under the translation of $p^n\Lambda$. Put $C(n) := |\Lambda(n)|$. Our goal is to compute $C(N)$. It is trivial that $C(0) = |\Lambda| = p^{(2m+1)N}$.

We first compute $C(1)$ as a start. The quotient $\mathbb{F}_p$-vector space $\Lambda/p\Lambda$, equipped with the $\mathbb{F}_p$-bilinear form $(\ ,\ ) \mod p$, becomes a nondegenerate quadratic space of dimension $2m+1$. It is clear that an element $x \in \Lambda/p\Lambda$ belongs to $\Lambda(1)/p\Lambda$ if and only if $x$ is isotropic.

We claim that for every nondegenerate quadratic space $V$ of dimension $2m+1$ over $\mathbb{F}_p$, the number of isotropic elements is $p^{2m}$. We prove by induction on $m$. When $m = 0$, it is trivial. For $m > 0$, we may pick two elements $v_1, v_2 \in V$ such that $(v_1, v_1) = 1$, $(v_2, v_2) = -1$, and $(v_1, v_2) = 0$. Let $V'$ be the orthogonal complement of $\{v_1, v_2\}$ in $V$, which is a nondegenerate quadratic space of dimension $2m-1$. For every $x \in V$, write $x = x_1v_1 + x_2v_2 + x'$ with $x_1, x_2 \in \mathbb{F}_p$ and $x' \in V'$ under the above orthogonal decomposition. Then $(x,x) = x_1^2 - x_2^2 + (x', x')$. To count the number of $x$ with $(x,x) = 0$, there are two cases: If $(x', x') = 0$, then the number of $(x_1, x_2) \in \mathbb{F}_p^2$ satisfying $x_1^2 - x_2^2 = 0$ is $2p-1$. If $(x', x') \neq 0$, then the number of $(x_1, x_2) \in \mathbb{F}_p^2$ satisfying $x_1^2 - x_2^2 = -(x', x') \neq 0$ is $p-1$ (as $p \neq 2$). Thus, by the induction hypotheses, the total number of $x$ that is isotropic is $p^{2m-2}(2p-1) + (p^{2m-1} - p^{2m-2})(p-1) = p^{2m}$.

After all, we get $C(1) = p^{2m}p^{(2m+1)(N-1)}$.

In general, we put $\Lambda(n)' := \Lambda(n) \setminus p\Lambda$ and $\Lambda(n)'' := \Lambda(n) \cap p\Lambda$ for $n \geq 1$, and define $C(n)'$ and $C(n)''$ to be the cardinality of $\Lambda(n)'$ and $\Lambda(n)''$ respectively. We claim

1. For $n \geq 2$, the map $\Lambda(n-2)/p^{n-1}\Lambda \to \Lambda(n)''/p^n\Lambda$ given by multiplying $p$ is a bijection.
2. For $n \geq 2$, the map $\Lambda(n)'/p^n\Lambda \to \Lambda(n-1)'/p^{n-1}\Lambda$ induced by the inclusion $\Lambda(n)' \to \Lambda(n-1)'$ is a $p^{2m}$-to-1 map.

We postpone the proofs of the two claims and compute $C(n)$ first. The first claim implies that $C(n)'' = p^{-(2m+1)}C(n-2)$ for $n \geq 2$; while the second claim implies that $C(n)' = p^{-1}C(n-1)'$ for $n \geq 2$. From the second one, we deduce

$$C(n)' = p^{-(n-1)}C(1)' = p^{-(n-1)}(C(1) - |p\Lambda|) = p^{(2m+1)(N-1)-(n-1)}(p^{2m} - 1).$$

Combining the first one, we get a recursive formula

$$C(n) = C(n)' + C(n)'' = p^{-(2m+1)}C(n-2) + p^{(2m+1)(N-1)-(n-1)}(p^{2m} - 1).$$

It follows easily that

$$C(N) = p^{(2m+1)r+2m(N-2r)} + \frac{p^{(2m-1)r} - 1}{p^{(2m-1)} - 1}p^{(2m+1)r-1+2m(N-2r)}(p^{2m} - 1)$$

where $r := \lfloor N/2 \rfloor$.

It remains to show the two claims.

The first one is easy by observing that an element $x \in \Lambda$ belongs to $\Lambda(n-2)$ if and only if $px$ belongs to $\Lambda(n)$.

For the second one, take an element $y \in \Lambda(n-1)'/p^{n-1}\Lambda$. Its preimage consists of elements $y + p^{n-1}x$ for $x \in \Lambda/p\Lambda$ satisfying

$$(y, y) + 2p^{n-1}(x, y) + p^{2n-2}(x, x) \in p^n\mathbb{Z}/p^N\mathbb{Z}.$$

As $n \geq 2$ and $(y, y) \in p^{n-1}\mathbb{Z}/p^N\mathbb{Z}$, the above equation amounts to

$$p^{1-n}(y, y) + 2(x, y) \equiv 0 \mod p.$$

Now since $p \neq 2$, $y \notin p\Lambda/p^{n-1}\Lambda$, and $\Lambda/p\Lambda$ is a nondegenerate quadratic space, the above equation is a polynomial (about $x$) of degree one on $\Lambda/p\Lambda$, hence whose solution set has cardinality $p^{2m}$.

## Question 2

Let $p > 2023$ be a prime number. Denote by $\mathcal{X}$ the set of all 2000-dimensional subspaces of the $\mathbb{F}_p$-vector space $\mathbb{F}_p^{2023}$. Find the minimal cardinality of a subset $\mathcal{Y}$ of $\mathcal{X}$ such that

$$\sum_{W \in \mathcal{Y}} (V \cap W) = V$$

holds for every $V \in \mathcal{X}$; justify your answer.

## Answer 2

The minimum cardinality is 25.

First we show that there exists $\mathcal{Y} \subset \mathcal{X}$ of cardinality 25 such that

$$\sum_{W \in \mathcal{Y}} (V \cap W) = V \tag{1}$$

holds for every $V \in \mathcal{X}$. Let $\{e_1, \ldots, e_{2023}\}$ be the standard basis of $\mathbb{F}_p^{2023}$. Put

$$W_k := \operatorname{span}\big\{\{e_1, \ldots, e_{2023}\} \setminus \{e_{23(k-1)+1}, \ldots, e_{23k}\}\big\}$$

for $1 \leqslant k \leqslant 25$. We show that $\mathcal{Y} := \{W_1, \ldots, W_{25}\}$ satisfies the requirement. Take $V \in \mathcal{X}$ and note that $\sum_{W \in \mathcal{Y}} (V \cap W) \subset V$. If (1) does not hold for $V$, then

$$d := \dim \sum_{W \in \mathcal{Y}} (V \cap W) \leqslant 1999.$$

In this case, for every $W \in \mathcal{Y}$, the codimension of $V \cap W$ in $\sum_{W \in \mathcal{Y}} (V \cap W)$ is at most $d - (2000 - 23) \leqslant 22$. It follows that

$$\dim \bigcap_{W \in \mathcal{Y}} (V \cap W) \geqslant 1999 - 25 \cdot 22 = 1449.$$

On the other hand, we have

$$\dim \bigcap_{W \in \mathcal{Y}} (V \cap W) \leqslant \dim \bigcap_{W \in \mathcal{Y}} W = 2023 - 25 \cdot 23 = 1448,$$

which is a contradiction.

It remains to show that for every subset $\mathcal{Y} \subset \mathcal{X}$ of cardinality at most 24, there exists $V \in \mathcal{X}$ for which (1) fails. Denote by $(\mathbb{F}_p^{2023})^\vee$ the $\mathbb{F}_p$-linear dual of $\mathbb{F}_p^{2023}$. For every $W \in \mathcal{Y}$, take a nonzero element $f_W \in (\mathbb{F}_p^{2023})^\vee$ such that $W \subset \operatorname{Ker} f_W$. Let $M$ be a 24-dimensional subspace of $(\mathbb{F}_p^{2023})^\vee$ containing $\{f_W \mid W \in \mathcal{Y}\}$. Since $p > 24$, there exists a 23-dimensional subspace $N$ of $M$ such that $N \cap \{f_W \mid W \in \mathcal{Y}\} = \emptyset$. Put $V := \bigcap_{f \in N} \operatorname{Ker} f$. Then $V$ belongs to $\mathcal{X}$ and satisfies $V \cap \operatorname{Ker} f_W = \bigcap_{f \in M} \operatorname{Ker} f$ for every $W \in \mathcal{Y}$. It follows that

$$\sum_{W \in \mathcal{Y}} (V \cap W) \subset \sum_{W \in \mathcal{Y}} (V \cap \operatorname{Ker} f_W) = \bigcap_{f \in M} \operatorname{Ker} f,$$

which is properly contained in $V$.

# Geometry & Topology

## Question 1

Let $M_n = \{(u, v) \in S^n \times S^n | u \cdot v = 0\}$, where $n \geq 2$, and $u \cdot v$ is the Euclidean inner product of $u$ and $v$. Suppose that the topology of $M_n$ is induced from $S^n \times S^n$.

(1) Prove that $M_n$ is a connected regular submanifold of $S^n \times S^n$;

(2) $M_n$ is a Lie Group if and only if $n = 2$.

## Answer 1

(1) Consider function $f : S^n \times S^n \to \mathbb{R}, f(u, v) = u \cdot v$. If $u_0 \neq \pm v_0$, we prove that the rank of $f$ at $(u_0, v_0)$ is 1. In fact, let $w_0' = v_0 - (u_0 \cdot v_0)u_0$, then

$$w_0' \cdot u_0 = 0, w_0' \cdot v_0 = 1 - (u_0 \cdot v_0)^2 > 0.$$

Let $w_0 = w_0'/\|w_0'\| \in S^n$, define

$$\sigma(t) = (u_0 \cos t + w_0 \sin t, v_0), t \in \mathbb{R}.$$

$\sigma$ is a curve on $S^n \times S^n$ starting from $(u_0, v_0)$. Since

$$f_{*(u_0,v_0)}\sigma'(0) = \frac{d}{dt}\bigg|_{t=0} f(\sigma(t)) = w_0 \cdot v_0 > 0,$$

the rank of $f$ at $(u_0, v_0)$ is 1. In particular, the rank of $f$ on $M_n = f^{-1}(0)$ is identically 1, thus $M_n$ is a regular submanifold of $S^n \times S^n$. As a differentiable manifold, $M_n$ is actually the unit sphere bundle of the tangent bundle of $S_n$, so it is connected.

(2) For $n = 2$, consider the map

$$\phi : M_2 \to SO(3), (u, v) \mapsto (u, v, u \times v).$$

$\phi$ is a diffeomorphism, thus induces the Lie group structure on $M_2$.

For $n = 3$, since $S^3$ has trivial tangent bundle, $M_3 = S^3 \times S^2$. In particular, $M_3$ is simply connected and has nontrivial $H^2(M_3; \mathbb{R})$, thus cannot be a Lie group by standard facts.

For $n = 4$, consider the Gysin sequence

$$\to H^{i-4}(S^4; \mathbb{R}) \overset{\cup e}{\to} H^i(S^4; \mathbb{R}) \to H^i(M^4; \mathbb{R}) \to H^{i-3}(S^4; \mathbb{R}) \overset{\cup e}{\to} H^{i+1}(S^4; \mathbb{R}) \to \cdots$$

Since the Euler class of $TS^4$ is nontrivial, it follows that $H^3(M_4; \mathbb{R})$ is trivial. Again, since $M_4$ is simply connected, it cannot be a Lie group by standard facts.

For $n \geq 5$, by Gysin sequence it is easy to see that $H^3(M_n; \mathbb{R})$ is trivial. By the same reason as above, $M_n$ cannot be a Lie group.

**Remark:** Any simply connected compact Lie Group must has trivial $H^2$ and nontrivial $H^3$.

## Question 2

Let $M$ be a complete Riemannian orbifold of dimension $n \geq 2$ with isolated singular points.

(1) Assume that the sectional curvature is nonnegative and the volume growth at infinity is positive, namely

$$\lim_{r\to\infty} \frac{vol(B(x,r))}{r^n} > 0. \tag{2}$$

at some $x$. Show that the number of singular points is at most 1.

(2) Assume that M is compact and the Ricci curvature is nonnegative. Assume in addition that the local group $\Gamma_x$ at some singular point $x$ acts irreducibly in the sense that it has no non-trivial invariant subspace. Show that $b_1(M) = 0$.

Here are some notions that may be used in the problem. An orbifold of dimension $n$ is a topological space $M$ such that at each $x$ there is a neighborhood $U_x$ and a homeomorphism $\phi_x : U_x \to B/\Gamma_x$ with $\phi_x(x) = 0$, where $B$ is the unit ball in the $n$-dimensional Euclidean space $\mathbb{R}^n$ and $\Gamma_x$ is a discrete subgroup of $O(n)$ acting linearly and effectively on $\mathbb{R}^n$. The data $(U_x, \Gamma_x, \phi_x)$ is called an orbifold chart at $x$. A point $x$ is called a regular point if $\Gamma_x$ is trivial in certain chart; otherwise, it is called a singular point. A Riemannian orbifold $M$, with isolated singular points $\Sigma = \{p_i\}_{i\in I}$, is an orbifold satisfying the followings: (a) $M\backslash\Sigma$ is a smooth manifold with a (non-complete) Riemannian metric $g$, (b) there is a family of charts $(U_i, \Gamma_i, \phi_i)$ at $p_i$ such that (b1) $\phi_i$ is smooth on $U_i\backslash\{p_i\}$ and (b2) the metric $g = \phi_i^*(\tilde{g}_i)$ where $\tilde{g}_i$ is a smooth, $\Gamma_i$-invariant metric on $B$. The Riemannian orbifold is called complete if the induced metric space is complete.

## Answer 2

(1) Suppose by contradiction that the orbifold points is bigger than or equal to 2, one picks two such points $P, Q$. Let $\gamma$ be a shortest geodesic connecting them. Then, any other point $x \in M$ achieves the distance $d(x,\gamma)$ at some $y \in \gamma$. The geodesic from $x$ to $y$ is perpendicular to $\gamma$ once $y$ is an interior point on $\gamma$, otherwise it makes an angle with $\gamma$ less than or equal to $\pi/2$ if $y = P$ or $Q$ (orbifold point property). In either case, one can apply the variation formula of volume on the tube neighborhood of $\gamma$. Precisely, by Heintze-Karcher volume comparison theorem for tubes,

$$vol(B(r,\gamma)) \leq L(\gamma)\cdot\omega_{n-2}\cdot r^{n-1} \tag{3}$$

where $\omega_{n-2}$ is the volume of unit $(n-2)$-sphere. In particular, the asymptotic volume at infinity vanishes. A contradiction.

(2) First of all, any element in $b_1(M)$ has a harmonic representation. Then, by Bochner formula to harmonic 1-form $\omega$,

$$\triangle|\omega|^2 = 2|\nabla\omega|^2 + 2Ric(\omega^\sharp, \omega^\sharp)$$

where $\omega^\sharp$ is the dual vector field of $\omega$. In particular, the maximum principle shows that $\omega$ is parallel.

If $b_1 \neq 0$, one pick a non-trivial parallel one form $\omega$. On the uniformization at the orbifold point $x$, the covector $\omega(x)$ is invariant by $\Gamma_x$, so it spans an invariant subspace of $\Gamma_x$. Contradiction.

# Analysis & Differential Equations

## Question 1

Let $P_1$, $P_2$, …, $P_n$ be $n$ points in $\mathbb{R}^d$.

(1) For $d = 1$, show that

$$\sum_{i<j} (-1)^{i+j} |P_i - P_j| \le 0.$$

(2) Prove that the above inequality holds for all $d \ge 1$. Here $|P_i - P_j|$ denotes the Euclidean distance of the two points $P_i$, $P_j$.

## Answer 1

For the case in dimension one, we prove by induction. The case when $n \le 2$ is trivial. Without loss of generality assume that $0 \le x_1 \le x_2 \le \ldots \le x_k$ stands for the points with odd index and $0 \le y_1 \le y_2 \le \ldots \le y_l$ represents the points with even index. In particular we have $l = k$ or $k - 1$. By induction assumption, it suffices to show that

$$\sum_{i<k} |x_k - x_i| + \sum_{j<l} |y_l - y_j| - \sum_{j<l} |x_k - y_j| - \sum_{i<k} |y_l - x_i| - |x_k - y_l| \le 0.$$

In fact

$$\begin{aligned}
&\sum_{j<l} |x_k - y_j| + \sum_{i<k} |y_l - x_i| + |x_k - y_l| \\
&\ge (l-1)x_k - \sum_{j<l} y_j + (k-1)y_l - \sum_{i<k} x_i + |x_k - y_l| \\
&= \sum_{i<k} |x_k - x_i| + \sum_{j<l} |y_l - y_j| + (l-k)(x_k - y_l) + |x_k - y_l| \\
&\ge \sum_{i<k} |x_k - x_i| + \sum_{j<l} |y_l - y_j|
\end{aligned}$$

as $|l - k| \le 1$.

For the higher dimensional case, we rely on the projection to the unit sphere and the reduce the problem to the case in dimension one. Let $\omega \in \mathbb{S}^{d-1}$. Let $P_i^\omega$ be the projection of the point $P_i$ to the line $O\omega$ passing through the origin and $\omega$. We claim that

$$\int_{\mathbb{S}^{d-1}} |P_i^\omega - P_j^\omega| d\omega = C|P_i - P_j|$$

for some constant $C$ independent of $i$, $j$. In fact, by spherical symmetry, we can assume that $P_i$, $P_j$ lines on the $x_1$ axis. By translation symmetry, we could move the origin to $P_i$. Finally by scaling symmetry, we see that the above integral is proportional to the length $|P_i - P_j|$ with a constant $C$ independent of $P_i$ and $P_j$. Therefore

$$\sum_{i<j}(-1)^{i+j}|P_i - P_j| = C^{-1}\sum_{i<j}\int_{\mathbb{S}^{d-1}}|P_i^\omega - P_j^\omega|d\omega \leq 0.$$

## Question 2

Let $f(z)$ be an analytic function defined on the complex plane $\mathbb{C}$. Define $f^{(n+1)}(z) = f(f^{(n)}(z))$ with $f^{(1)}(z) = f(z)$. Is there analytic function $f(z)$ on $\mathbb{C}$ such that $f^{(2023)}(z) = e^{2023z}$ for all $z \in \mathbb{C}$? Prove your assersion.

## Answer 2

The answer is no. Otherwise assume $f$ is analytic verifying the equation $f^{(2023)}(z) = e^{2023z}$. In particular we can conclude that the image $f(\mathbb{C})$ of $f$ must be $C^* = \mathbb{C}\backslash\{0\}$. Hence there exists an entire function $g(z)$ on $\mathbb{C}$ such that $f(z) = e^{g(z)}$. By the assumption there is an integer $k$ such that

$$g(f^{(2022)}(z)) = 2023z + 2k\pi i, \quad \forall z \in \mathbb{C}.$$

Replacing $z$ with $f(z)$, we have $g(e^{2023z}) = 2023f(z) + 2k\pi i$, which implies that

$$2023f(\mathbb{C}) + 2k\pi i = \mathbb{C}\backslash\{2k\pi i\} = g(e^{2023\mathbb{C}}) = \mathbb{C}\backslash\{g(0)\}.$$

This means that $g(0) = 2k\pi i$ and

$$g(f^{(2022)}(0)) = 2k\pi i = g(0)$$

Note that $f$ is injective and we can require that $g$ is also injective. We thus conclude that $f^{(2022)}(0) = 0$, contradicting to $f(\mathbb{C}) = C^*$.

# Applied & Computational Mathematics

## Question 1

In order to align a large language model (LLM) with human preferences, it is necessary to fine-tune the LLM to adhere to rankings derived from human feedback. Given a specific prompt, let the LLM generate $n$ responses. A human labeler will then rank these $n$ responses from best to worst. Denote the ranking by $\pi$ (for example, the $\pi(1)$-th response is considered the best). Suppose we have a function $G$ that is strongly concave (so $-G$ is strongly convex) and increasing on the interval $[-1, 1]$. We aim to train a reward model that assigns a rating $0 \le r_i \le 1$ to the $i$-th response. Ideally, the rewards $r_i$ should be the solution to the following optimization program:

$$\max_{0 \le r_1, \ldots, r_n \le 1} \sum_{i<j} G(r_{\pi(i)} - r_{\pi(j)}).$$

(1) Explain why

$$L(r_1, \ldots, r_n) := \sum_{i<j} G(r_{\pi(i)} - r_{\pi(j)})$$

is concave but not strongly concave (you can assume sufficient smoothness of $G$ if it is helpful). Prove that, despite this, the optimization program stated above has a *unique* solution $(r_1^\star, \ldots, r_n^\star) \in [0, 1]^n$.

(2) Prove that the solution must satisfy

$$1 = r^\star_{\pi(1)} \ge r^\star_{\pi(2)} \ge \cdots \ge r^\star_{\pi(n)} = 0$$

and $r^\star_{\pi(i)} + r^\star_{\pi(n+1-i)} = 1$ for all $i = 1, \ldots, n$.

(3) Now, consider the limiting version of the problem above. Assume that as $n \to \infty$, the empirical distribution of $r_1^\star, \ldots, r_n^\star$ converges to a probability measure $\mu$ on the interval $[0, 1]$. The problem can then be reformulated as

$$\sup_\mu \mathbb{E}_{X, X' \overset{iid}{\sim} \mu} G(|X - X'|),$$

where $X, X'$ are independent and identically distributed draws from $\mu$. If a probability measure $\mu^\star$ maximizes $\mathbb{E}_{X, X' \overset{iid}{\sim} \mu} G(|X - X'|)$, prove that $\mathbb{E}_{X \sim \mu^\star} G(|X - c|)$ is independent of $c \in [0, 1]$.

## Answer 1

Without loss of generality, assume $\pi(i) = i$. Let $H$ be the Hessian of $L$. Then

$$H_{ij} = -G''(r_{\min(i,j)} - r_{\max(i,j)})$$

if $i \ne j$ and

$$H_{ii} = -\sum_{j \ne i} H(ij).$$

Hence, $H$ is diagonal dominant. $H$ is then positive semi-definite. The only eigenvector with eigenvalue 0 is the all ones vector. This shows that the only null direction is translation. However, since $G$ is increasing, $r_1^\star = 1$ and $r_n^\star = 0$. Thus it is unique and (1) is proved.

Due to the monotonicity of $G$, it is clear that the optimal solution must have the largest gap $r^\star_{\pi(1)} - r^\star_{\pi(n)}$. Hence, $r^\star_{\pi(1)} = 1$ and $r^\star_{\pi(n)} = 0$. Moreover, if $r^\star$ is a solution, then must be $1 - r^\star$ (reverse the ordering). This proves (2).

Use calculus of variations. For two probability measure $\mu$ and $\nu$ on $\mathcal{B}([0,1])$, $(\mu-\nu)([0,1]) = 0$. Suppose $\mu$ satisfies $\mathbb{E}_{X\sim\mu} G(|X-c|) = K$ does not depend on $c \in [0,1]$. Note that

$$\langle \nu - \mu, \mu \rangle = \int_{[0,1]} \left( \int_{[0,1]} |x-y|^\beta \mu(dx) \right) (\nu - \mu)(dy) = \int_{[0,1]} K(\nu-\mu)(dy) = 0.$$

Next, $\langle \nu - \mu, \nu - \mu \rangle \leq 0$. Therefore,

$$\langle \nu, \nu \rangle = \langle \mu, \mu \rangle + 2 \langle \nu - \mu, \mu \rangle + \langle \nu - \mu, \nu - \mu \rangle \leq \langle \mu, \mu \rangle .$$

This proves (3). However, note that points should deducted if the student does not consider the situation where the support of $\mu$ is strictly smaller than $[0,1]$.

## Question 2

Consider a linear Fokker-Planck equation as

$$\partial_t f = \Delta f + \nabla \cdot [f \nabla V(x)], \quad t > 0,\ x \in \Omega \subset \mathbb{R}^d, \tag{4}$$

where $f = f(t,x) \geq 0$ is the probability density function of time $t$ and position $x$, and $V(x)$ is a given function which has a lower bound in $\Omega$. Written in form (4), the Fokker-Planck equation is just a convection-diffusion equation. To numerically solve (4), we define $M(x) = \exp(-V(x))$ and use an equivalent form of (4):

$$\partial_t f = \nabla \cdot \left( M \nabla \left( \frac{f}{M} \right) \right). \tag{5}$$

In this way, the equation can be viewed as a variable coefficient diffusion equation and can be naturally discretized by standard finite difference scheme. Specifically, we assume $d = 1$, $\Omega = [a,b]$ and partition $[a,b]$ uniformly into $N$ cells with size $\Delta x = (a-b)/N$, cell center $x_i = a + (i - 1/2)\Delta x$, $i = 1, \cdots, N$, and cell interface $x_{i+1/2} = a + i\Delta x$, $i = 0, \cdots, N$. Applying central finite difference in space and forward Euler in time to (5) yields

$$\frac{f_i^{n+1} - f_i^n}{\Delta t} = \frac{F^n_{i+1/2} - F^n_{i-1/2}}{\Delta x}, \quad F^n_{i+1/2} := \frac{1}{\Delta x} M_{i+1/2} \left( \frac{f^n_{i+1}}{M_{i+1}} - \frac{f_i^n}{M_i} \right), \tag{6}$$

where $f_i^n$ is the numerical approximation of $f$ at $x_i$ and time $t_n = n\Delta t$, $M_i$ and $M_{i+1/2}$ are the values of $M$ at $x_i$ and $x_{i+1/2}$, respectively. At the boundary $a$ and $b$, we take $F^n_{1/2} = F^n_{N+1/2} = 0$ to account for the no-flux boundary condition.

a) For the scheme (6), prove the energy functional $E(f) := \int_\Omega f \ln\left(\frac{f}{M}\right) \mathrm{d}x$ decays at the discrete level, that is,

$$E_\Delta(f_i^{n+1}) \le E_\Delta(f_i^n), \quad \text{where} \;\; E_\Delta(f_i^n) := \sum_{i=1}^N f_i^n \ln\left(\frac{f_i^n}{M_i}\right). \tag{7}$$

State clearly the conditions needed. Does the numerical solution converge to a steady state? If yes, find its form. If not, explain it.

b) Alternatively, one can apply backward Euler in time to obtain

$$\frac{f_i^{n+1} - f_i^n}{\Delta t} = \frac{F_{i+1/2}^{n+1} - F_{i-1/2}^{n+1}}{\Delta x}.$$

For this scheme, repeat the same questions as in part a).

## Answer 2

a) The explicit scheme can be written equivalently as

$$\frac{f_i^{n+1}}{M_i} = \left(1 - \frac{\Delta t}{\Delta x^2}\frac{M_{i+1/2}}{M_i} - \frac{\Delta t}{\Delta x^2}\frac{M_{i-1/2}}{M_i}\right)\frac{f_i^n}{M_i} + \frac{\Delta t}{\Delta x^2}\frac{M_{i+1/2}}{M_i}\frac{f_{i+1}^n}{M_{i+1}} + \frac{\Delta t}{\Delta x^2}\frac{M_{i-1/2}}{M_i}\frac{f_{i-1}^n}{M_{i-1}}. \tag{8}$$

Define $g_i^n = f_i^n/M_i$, $\mathbf{g}^n = (g_1^n, \dots, g_N^n)^T$, then (8) can be written as

$$\mathbf{g}^{n+1} = A\mathbf{g}^n, \tag{9}$$

where $A$ is a tridiagonal matrix with entries

$$a_{ii} = 1 - \frac{\Delta t}{\Delta x^2}\frac{M_{i+1/2}}{M_i} - \frac{\Delta t}{\Delta x^2}\frac{M_{i-1/2}}{M_i}, \quad a_{i,i-1} = \frac{\Delta t}{\Delta x^2}\frac{M_{i-1/2}}{M_i}, \quad a_{i,i+1} = \frac{\Delta t}{\Delta x^2}\frac{M_{i+1/2}}{M_i}. \tag{10}$$

For the matrix $A$, one can verify the following properties:

1. By choosing $\Delta t \le \Delta x^2 \min_i \left\{\frac{M_i}{M_{i+1/2} + M_{i-1/2}}\right\}$, one can make $A \ge 0$, i.e., every entry of $A$ is non-negative;
2. $A\mathbf{1} = \mathbf{1}$ ($\mathbf{1}$ is all one column vector), i.e., $\sum_j a_{ij} = 1$;
3. $A^T\mathbf{M} = \mathbf{M}$, where $\mathbf{M} = (M_1, \dots, M_N)^T$, i.e., $\sum_j M_j a_{ji} = M_i$.

Property 1 implies that the scheme is positivity-preserving so one can make sense of the entropy:

$$\begin{aligned} &f_i^n \ge 0 \text{ for all } i \Longrightarrow g_i^n = f_i^n/M_i \ge 0 \text{ for all } i \\ \Longrightarrow &g_i^{n+1} = \sum_j a_{ij} g_j^n \ge 0 \text{ for all } i \Longrightarrow f_i^{n+1} = M_i g_i^{n+1} \ge 0 \text{ for all } i. \end{aligned} \tag{11}$$

Properties 1 and 2 together imply that $g_i^{n+1} = \sum_j a_{ij} g_j^n$ is a convex combination so one can apply Jensen's inequality (see below).

Property 3 implies that the scheme conserves mass:

$$\sum_i f_i^{n+1} = \sum_i M_i g_i^{n+1} = \sum_i M_i \left( \sum_j a_{ij} g_j^n \right) = \sum_j \left( \sum_i M_i a_{ij} \right) g_j^n = \sum_j M_j g_j^n = \sum_j f_j^n. \tag{12}$$

For the discrete entropy, we have

$$\begin{aligned} E_\Delta(f_i^{n+1}) &= \sum_i f_i^{n+1} \ln \left( \frac{f_i^{n+1}}{M_i} \right) = \sum_i M_i g_i^{n+1} \ln g_i^{n+1} \leq \sum_i M_i \left( \sum_j a_{ij} g_j^n \ln g_j^n \right) \\ &= \sum_j \left( \sum_i M_i a_{ij} \right) g_j^n \log g_j^n = \sum_j M_j g_j^n \ln g_j^n = \sum_j f_j^n \ln \left( \frac{f_j^n}{M_j} \right) = E_\Delta(f_i^n), \end{aligned} \tag{13}$$

where we used Jensen's inequality to obtain the inequality. So the scheme also decays the entropy.

Furthermore, $E_\Delta(f_i^n)$ has a lower bound hence the sequence $\{E_\Delta(f_i^n)\}$ has a limit. Taking the limit on both sides of (13) yields $g_1^\infty = \cdots = g_N^\infty$. Using the mass conservation and form of $M$, this implies the steady state takes the form:

$$f_i^\infty = C \exp(-V(x_i)), \quad \text{with } C = \frac{\sum_i f_i^0}{\sum_i \exp(-V(x_i))} \text{ and } f_i^0 \text{ is the initial condition.} \tag{14}$$

b) The implicit scheme can be written equivalently as

$$\left( 1 + \frac{\Delta t}{\Delta x^2} \frac{M_{i+1/2}}{M_i} + \frac{\Delta t}{\Delta x^2} \frac{M_{i-1/2}}{M_i} \right) \frac{f_i^{n+1}}{M_i} - \frac{\Delta t}{\Delta x^2} \frac{M_{i+1/2}}{M_i} \frac{f_{i+1}^{n+1}}{M_{i+1}} - \frac{\Delta t}{\Delta x^2} \frac{M_{i-1/2}}{M_i} \frac{f_{i-1}^{n+1}}{M_{i-1}} = \frac{f_i^n}{M_i}. \tag{15}$$

Define $g_i^n = f_i^n / M_i$, $\mathbf{g}^n = (g_1^n, \ldots, g_N^n)^T$, then (15) can be written as

$$B \mathbf{g}^{n+1} = \mathbf{g}^n, \tag{16}$$

where $B$ is a tridiagonal matrix with entries

$$b_{ii} = 1 + \frac{\Delta t}{\Delta x^2} \frac{M_{i+1/2}}{M_i} + \frac{\Delta t}{\Delta x^2} \frac{M_{i-1/2}}{M_i}, \quad b_{i,i-1} = -\frac{\Delta t}{\Delta x^2} \frac{M_{i-1/2}}{M_i}, \quad b_{i,i+1} = -\frac{\Delta t}{\Delta x^2} \frac{M_{i+1/2}}{M_i}. \tag{17}$$

For the matrix $B$, one can verify the following properties:

1. $B$ is a M-matrix, then $B^{-1} \geq 0$, i.e., every entry of $B^{-1}$ is non-negative;
2. $B\mathbf{1} = \mathbf{1}$, then $B^{-1}\mathbf{1} = \mathbf{1}$;
3. $B^T \mathbf{M} = \mathbf{M}$, then $(B^{-1})^T \mathbf{M} = \mathbf{M}$.

Now let's view the scheme (16) as

$$\mathbf{g}^{n+1} = B^{-1}\mathbf{g}^n. \tag{18}$$

For the matrix $B^{-1}$, it satisfies exactly the same properties as the matrix $A$ in the explicit scheme. Therefore, positivity, mass conservation, entropy dissipation, and steady state can be readily obtained.

# Combinatorics & Probability

## Question 1

In a dance party initially there are 20 girls and 22 boys in the pool and infinitely many more girls and boys waiting outside. In each round, a participant is picked uniformly at random; if a girl is picked, then she invites a boy from the pool to dance and then both of them leave the party after the dance; while if a boy is picked, then he invites a girl and a boy from the waiting line and dance together. The three of them all stay after the dance. The party is over when there are only (two) boys left in the pool.

(1) What is the probability that the party never ends?

(2) Now the organizer of this party decides to reverse the rule, namely that if a girl is picked, then she invites a boy and a girl from the waiting line to dance and the three stay after the dance; while if a boy is picked, he invites a girl from the pool to dance and both leave after the dance. Still the party is over when there are only (two) boys left in the pool. What is the expected number of rounds until the party ends?

## Answer 1

(1) We consider a party with $n$ girls and $(n+2)$ boys. Let $X_k$ be the number of girls after $k$ rounds. Let $e_j$ be the probability that $(X_n)_{n\geq 0}$ reaches zero in a finite time when $X_0 = j$. By definition

$$e_j = p_j e_{j+1} + q_j e_{j-1},\ j \geq 1,\ \text{with } p_j = \tfrac{j+2}{2(j+1)},\ q_j = \tfrac{j}{2(j+1)} \text{ and } e_0 = 1.$$

By a classical result for the first passage time for Markov chains, $e_j$ has to be the minimal non-negative solution of the equation system. Solving this equation system recursively, we have

$$e_j = 1 - (1-e_1)\Big[1 + \frac{q_1}{p_1} + \cdots + \frac{q_1 q_2 \cdots q_{j-1}}{p_1 p_2 \cdots p_{j-1}}\Big].$$

It is easy to see that the term in the big bracket on the RHS is equal to $2 - \frac{2}{j+1}$. Hence the minimal solution is attained at $e_j = 1/(j+1)$. Therefore the party never ends with probability 20/21.

(2) Let $d_j$ be the expected time until $(j-1)$ girls remain, remain, having started with $j$ girls and $(j+2)$ boys. Similarly, we have

$$d_j = 1 + q_j[d_{j+1} + d_j].$$

Letting $f_j = d_j + (2j+1)$, we simplify the equation above as

$$(j+2)f_j = j f_{j+1}, \quad \text{hence } f_j = \frac{j(j+1)}{2} f_1.$$

The expected rounds of the party is

$$s_n := \sum_{i=1}^{n} d_j = n(n+2) + f_1 \sum_{j=1}^{n} \frac{j(j+1)}{2}.$$

Again by the classical property of mean first passage time of Markov chain, $s_n$, $n \geq 0$, has to be the minimal non-negative solution of this equation system with $s_0 = 0$. This is attained when $f_1 = 0$, and hence $s_{20} = 420$.

## Question 2

Given positive integers $k \geq 2$ and $m$ sufficiently large. Let $\mathcal{F}_m$ be the infinite family of all the (not necessarily square) binary matrices which contain exactly $m$ 1's. Denote by $f(m)$ the maximum integer $L$ such that for every matrix $A \in \mathcal{F}_m$, there always exists a binary matrix $B$ of the same dimension such that (1) $B$ has at least $L$ 1-entries; (2) every entry of $B$ is less or equal to the corresponding entry of $A$; (3) $B$ does not contain any $k \times k$ all-1 submatrix. Show the equality

$$\lim_{m \to \infty} \frac{\ln f(m)}{\ln m} = \frac{k}{k+1}.$$

## Answer 2

The problem can be reformulated in graph theory language: find the largest $f(m)$ such that every bipartite graph $G$ with $m$ edges contains a $K_{k,k}$-free subgraph with at least $f(m)$ edges. We claim that $f(m) = \Theta(m^{\frac{k}{k+1}})$. This gives

$$\lim_{m \to \infty} \frac{\ln f(m)}{\ln m} = \frac{k}{k+1}.$$

To show $f(m) = \Omega(m^{\frac{k}{k+1}})$, we take every edge of $G$ independently with probability $p \in (0, 1)$ and denote by $G'$ the resulted random subgraph. Define two random variable $X = e(G')$, the number of edges in $G'$, $Y$ which counts the number of copies of $K_{k,k}$ in $G'$. Note that the number of copies of $K_{k,k}$ does not exceed $m^k$. Using linearity of expectation,

$$\mathbb{E}[X] = pm,$$

$$\mathbb{E}[Y] \leq p^{k^2} m^k.$$

Let $Z = X - Y$, then by removing an edge from each of the $Y$ copy of $K_{k,k}$ in $G'$, we obtain a graph $G''$ that is $K_{k,k}$-free and contains at least $X - Y$ edges. Therefore, there exists such $G''$ with at least $\mathbb{E}[Z] \geq pm - p^{k^2} m^k$ edges. We let $p = \frac{1}{k^{2/(k^2-1)} m^{1/(k+1)}}$ to maximize the right hand side. Calculations give

$$e(G'') \geq \mathbb{E}[Z] = \Omega(m^{\frac{k}{k+1}}).$$

To see $f(m) = O(m^{\frac{k}{k+1}})$, for sufficiently large $m$, let $G$ be the complete bipartite graph on $A \cup B$, with $|A||B| = m$. Suppose $H$ is a $K_{k,k}$-free subgraph of $G$. Then the number of copies of $K_{1,k}$ with one vertex in $A$ and $k$ in $B$ is equal to (let $d_v$ be the degree of $v$ in $H$)

$$\sum_{v \in A} \binom{d_v}{k} \geq |A| \binom{m/|A|}{k}.$$

Here we use the convexity of $\binom{x}{k}$. Suppose for every set $T \subset B$, $|T| = k$, $d_T$ is the number of vertices in $A$ adjacent to all of $T$ in $H$. Then by double counting

$$\sum_{T:T\subset B,|T|=k} d_T = \sum_{v\in A} \binom{d_v}{k} \geq |A|\binom{e(H)/|A|}{k},$$

and the number of copies of $K_{k,k}$ in $H$ is equal to

$$\sum_{T:T\subset B,|T|=k} \binom{d_T}{k}.$$

Since $H$ is $K_{k,k}$-free, we have $d_T \leq k-1$ for all $T$, therefore using the previous inequality

$$|A|\binom{e(H)/|A|}{k} \leq \binom{|B|}{k}(k-1),$$

and we have

$$e(H) = O(|B||A|^{(k-1)/k}).$$

Let $|A| = m^{k/(k+1)}$ and $|B| = m^{1/(k+1)}$, this inequality give $e(H) = O(m^{k/(k+1)})$.

**Remark:** This result was established by Conlon, Fox and Sudakov in [Short proofs of some extremal results II, *J. Combin. Theory Ser. B* **121** (2016), 173–196.].