Example Problems that ML tools should be better at

Semon Rezchikov

April 2025

Abstract

This document describes a toy problem that popped up in my research, which is solvable but should maybe be solved by machine; and a beautiful problem essentially in pure computer algebra which has probably profound mathematical significance.

1 Toy Problem

The Cauchy Riemann equation for maps $u : \mathbb{C} \to \mathbb{C}^n$ can be written as

$$\partial_1 u + J\partial_2 u = 0 \tag{1}$$

where the coordinates on $\mathbb{C}$ are $z = x_1 + ix_2$, and J is the operator of multiplication by i (so it is just a series of copies of the matrix for i on the diagonal).

Instead of having the target by $\mathbb{C}^n$, it turns out that we can have the target be a symplectic manifold M with a 'compatible' almost complex structure J; then, the same equation describes complex curves in this sympelctic manifold, and the analysis of this nonlinear analog of the Cauchy-Riemann equation is the foundation of almost all progress in symplectic geometry.

I've been studying a quaternionic analog of these equations, which turns out to appear naturally in a variety of mathematical situations. We can replace the domain $\mathbb{C}$ with the quaternions $\mathbb{H}$ (with coordinates $q = x_0 + ix_q + jx_2 + kx_3$) and consider maps $U : \mathbb{H} \to \mathbb{H}^n$, or more generally, to a manifold M equipped with a triple of endomorphisms of the tangent bundle I, J, K satisfying the quaternionic relations $I^2 = J^2 = K^2 = IJK = -1$; the corresponding equation is

$$\partial_0 U - I\partial_1 U - J\partial_2 U - K\partial_3 U = 0. \tag{2}$$

There is a $3D$ analog, where one considers maps $U : Im\mathbb{H} \to \mathbb{H}^n$, with corresponding equation

$$I\partial_1 U + J\partial_2 U + K\partial_3 U = 0. \tag{3}$$

I expect a completely systematic analogy between known results for (1) and (2).

In particular, there is a notion of a J-convex function

$$\rho : \mathbb{C}^n \to \mathbb{R}, \text{ or more generally } \rho : M \to \mathbb{R} \text{ when } M \text{ has a } J.$$

This is a very useful tool; I'll focus on the target being $\mathbb{C}^n$ for the time being, because the greater generality of the target being a manifold introduces only one additional complication which I'd rather avoid for the time being. The condition is that the differential 2-form

$$\sigma_J^\rho = -d(d\rho \circ J)$$

evaluates positively on every complex line in every tangent space to $\mathbb{C}^n$. This is an explicit linear algebra condition on the hessian of ρ at every point, and it turns out to imply that for every solution u to (1), the composition $\rho \circ u$ has no local maxima, which is very helpful for controlling the behavior of holomorphic curves (by confining them to certain regions defined by ρ).

My collaborator and I invented nice analogs of convexity for solutions to (2) and (3). The toy problems that came up are - how are these notions related to one another? I'll explain the notions and the questions below; we found a nice solution to one of them after some thinking, and haven't gotten around to solving the second one.

We call a function $\rho : \mathbb{H}^n \to \mathbb{R}$ *IJK-convex* if for every linear map

$$A : Im\mathbb{H} \to \mathbb{H}^n \text{ satisfying } IAi + JAj + KAk = 0$$

and any $x \in \mathbb{H}^n$, we have that, writing

$$\bar{A} = (id, A) : Im\mathbb{H} \to Im\mathbb{H} \times \mathbb{H}^n,$$

the pullback of the differential three-form

$$\Omega_x^\rho = dx_1 \wedge (\sigma_I^\rho)_x + dx_2 \wedge (\sigma_J^\rho)_x + dx_3 \wedge (\sigma_K^\rho)_x$$

by $\bar{A}$ is a negative multiple of the volume form $dx_1 \wedge dx_2 \wedge dx_3$ on $Im\mathbb{H}$.

We call a function $\rho : \mathbb{H}^n \to \mathbb{R}$ $\mathbb{H}$-convex if for every linear map

$$A : \mathbb{H} \to \mathbb{H}^n \text{ satisfying } A - IAi + JAj + KAk = 0$$

and any $x \in \mathbb{H}^n$, we have that, writing

$$\bar{A} = (id, A) : \mathbb{H} \to \mathbb{H} \times \mathbb{H}^n,$$

the pullback of the differential four-form

$$\bar{\Omega}_x^\rho = \omega_1 \wedge (\sigma_I^\rho)_x + \omega_2 \wedge (\sigma_J^\rho)_x + \omega_3 \wedge (\sigma_K^\rho)_x$$

under $\bar{A}$, where the two-forms ω_i are

$$\omega_1 = dx_0 \wedge dx_1 + dx_2 \wedge dx_3$$

$$\omega_2 = dx_0 \wedge dx_2 + dx_3 \wedge dx_1$$

$$\omega_3 = dx_0 \wedge dx_3 + dx_1 \wedge dx_2,$$

is a negative multiple of the volume form, i.e.

$$\bar{A}^* \bar{\Omega}^\rho_x = c dx_0 dx_1 dx_2 dx_3, c < 0$$

for every x and every A as above.

I will not go onto the theory of such functions, but I will point out that $\mathbb{H}$-convex functions are IJK-convex, which are in turn J-convex, in the following strong sense.

First, for every

$$\tau = (a, b, c) \in \mathbb{R}^3 \text{ such that } a^2 + b^2 + c^2 = 1, \text{ i.e. } \tau \in S^2$$

the operator

$$J_\tau = aI + bJ + cK \text{ satisfies } J_\tau^2 = -1,$$

and the symmetries of the equation (3) immediately imply

Lemma 1. *An IJK-convex function is J_τ-convex for every $\tau \in S^2$.*

There is a similar phenomenon with $\mathbb{H}$-convex functions. The point is that, for every

$$v \in \mathbb{H}, \|v\| = 1, \text{ i.e. } v \in S^3$$

we can define

$$i_v = viv^{-1}, j_v = vjv^{-1}, k_v = vkv^{-1};$$

we will have that $i_v, j_v, k_v \in Im\mathbb{H}$, and then, writing

$$i_v = a_i i + b_i j + c_i k, \tau_i = (a_i, b_i, c_i)$$

$$j_v = a_j i + b_j j + c_j k, \tau_j = (a_j, b_j, c_j)$$

$$i_k = a_k i + b_k j + c_k k, \tau_k = (a_j, b_j, c_j)$$

we can define

$$I_v = J_{\tau_i}, J_v = J_{\tau_j}, K_v = J_{\tau_k},$$

and then it is elementary to see

Lemma 2. *If ρ is $\mathbb{H}$-convex then it is (I_v, J_v, K_v)-convex for all $v \in S^3$.*

The question was:

Question 1. *Is the converse to Lemma 1 true? Is the converse to Lemma 2 true?*

After some real thinking, we figured out that the converse to Lemma 1 was false (and thus IJK-convexity cannot be reduced to J-convexity), and we never got around to investigating the converse to 2.

2 A basic fact about linear algebra that is wholly unproven

Let us sample a random $n \times n$ orthogonal matrix O in the natural way, namely from the Haar measure on the group $SO(n)$ of orthogonal matrices. This is the unique probability measure on $SO(n)$ such that the distribution of O and of O_2O are the same for any other orthogonal matrix O_2. Explicitly, this is sampled as follows: Sample an IID unit normal matrix K, and apply Gram-Schmidt to K to produce O.

Let us take an $n \times n$ matrix $A_0 = diag(\lambda_1, \ldots, \lambda_n)$ with $\lambda_1 < \lambda_2 < \ldots < \lambda_n$ a sequence of real numbers, and define the random symmetric matrix

$$A = OAO^{-T}.$$

This matrix A still has the eigenvalues $\lambda_1, \ldots, \lambda_n$ but is no longer diagonal. Let the upper-left $k \times k$ block submatrix of A be A^k, and let $\lambda_1^k \leq \ldots \leq \lambda_k^k$; these are all random variables. It is a surprising by well known and easy to establish fact that these numbers interlace, i.e.

$$\lambda_j^{k+1} \leq \lambda_j^k \leq \lambda_{j+1}^{k+1}$$

for every k, j where the above quantities have been defined. The set of quantities $\{\lambda_j^k\}_{1 \leq k < n, 1 \leq j \leq k}$ satisfying the inequalities above is the *Gelfand-Tsetlin polytope* $GT(\lambda_1, \ldots, \lambda_n)$ (of type A). Let $\rho(\vec{\lambda})$ be the joint probability density of these random variables; this turns out to be a smooth function.

Conjecture: The probability density ρ is log-concave, i.e. the hessian of $\log \rho$

$$H(\vec{\lambda})_{k_1 j_1, k_2 j_2} = \partial_{\lambda_{j_1}^{k_1}} \partial_{\lambda_{j_2}^{k_2}} \log \rho(\vec{\lambda})$$

is negative definite (where we think of the tuple (k_a, j_a) as a single index).

This is a completely concrete problem in symbolic algebra.

What is known:

A basic fact is that *one can write an explicit expression for* ρ; it will take me a bit to look it up, and from the perspective of computer tools for mathematicians, it is reasonable to ask whether a computer would be able to discover this fact as well. I believe that Holden Lee checked the conjecture above in the 3×3 case; it is essentially a hard IMO-type inequality problem. I think we also checked this numerically in higher dimensions, but my confidence is not 100 percent; neither I am completely sure at this moment if the sign of the hessian of $\log \rho$ is supposed to positive or negative, i.e. maybe the correct conjecture is actually the negative of the above. In general, many fascinating inequalities related to determinants of Cauchy matrices with values that are interlacing, seem like they are at the heart of the problem.

The reason for my confusion about the sign is as follows: one can replace the Haar-random orthogonal matrix O by a Haar-random *compact quaternionic* *matrix* V, i.e. a random $n \times n$ matrix V with quaternionic entries satisfying the equation

$$V^\dagger V = Id,\ (V^\dagger)_{ij} = \bar{V}_{ji},\ q = a + bi + cj + dk \leftrightarrow \bar{q} = a - bi - cj - dk.$$

It turns out that symmetric quaternionic matrices also have real eigenvalues, so one can proceed *exactly as before*, and define a new probability dentity function $\rho_{\mathbb{H}}(\vec{\lambda})$ on the same Gelfand-Tsetlin polytope. The explicit form of $\rho_{\mathbb{H}}$ shows that ρ *is log-concave if and only if* $\rho_{\mathbb{H}}$ *is log convex, and vice versa.*

Context:

a) The above problem appeared in the context of efficient efficient differentially private matrix factorization.

b) This is a mild-seeming perturbation around a story that has led to a lot of highly influential mathematics; but in this setting, nothing is known. Here is the known story.

Let us replace the Haar-random orthogonal matrix O by a Haar-random unitary matrix U. Then it is a basic fact, probably due to Konstant, that the corresponding density $\rho_{\mathbb{C}}$ is *constant on the Gelfand-Tsetlin polytope.* The essential reason why we can prove this is that matrices of the form UAU^{-1} are Hermitian, so matrices of the form $iUAU^{-1} = U(iA)U^{-1}$ are *skew-Hermitian*, i.e. they are in the *Lie algebra of the unitary group.* The set of skew-hermitian matrices conjugate to iA, namely $\mathcal{O}(\lambda_1, \ldots, \lambda_n)$, turn out then to be a *symplectic manifold*; moreover, the functions λ_j^k on $\mathcal{O}(\lambda_1, \ldots, \lambda_n)$ turn out to be the *Hamiltonian functions* for a *completely integrable system* (called the Gelfand-Tsetlin system) on this symplectic manifold. It is then a completely general theorem of Duistermaat-Heckman that *the pushforward of the volume form on a symplectic manifold* M *along a function* $\mathcal{H} : M \to \mathbb{R}^N$ *whose coordinates are a Poisson-commuting family of functions on* M *is always log-convex*, and moreover the proof actually shows you that if we can take $N = \dim M/2$, i.e. the system is completely integrable, then the pushforward of the density is in fact constant on its support.

The discovery of the Duistermaat-Heckman (DH) thereoem in fact postdated the knowledge of the constancy of $\rho_{\mathbb{C}}$, and the latter discovery motivated the proof of the DH theorem, which in turn has many, many applications. IN particular, there is a version of the Gelfand-Tselin system, the Gelfand Tsetlin polytope, and this convexity result, for any lie group G. *However, these latter results do not help to understand our problem* for the following reason: they tell us something about eigenvalues of the ensemble of matrices

$$OBO^{-1} \text{ for } B \text{ a given } \textbf{skew-symmetric matrix } B \in Lie(SO(n))$$

an unlike in the unitary case, where the Lie algebra of the unitary group is identified with Hermitian matrices under multiplication by i, which is an identification that commutes with the group action, there is no similar identification between symmetric and skew-symmetric real matrices – these are spaces of different dimensions! All other aspects of the parallel break down as well – the space of symmetric matrices with fixed eigenvalues is not a symplectic manifold as it is not even dimensional, and so forth. Yet, we have evidence to the conjecture described above that some aspect of this more general story should persist in this setting; and speaking more broadly, understanding the conjecture above would likely lead to new structures of great value, just as the corresponding story in the unitary case did historically.