# Research Statement - Kevin Shu

Computer optimization plays a central role in society as it becomes increasingly capable of providing solutions to problems beyond what humans can understand. In many applications of optimization, it is not sufficient to produce a good locally optimal solution to a problem; it is also necessary to be able to confidently assert that no better solution exists. In applications such as robotics or industrial engineering, it is often necessary to guarantee that a system does not exceed some critical capacity restriction, which can also be viewed as asserting a global optimality guarantee for some optimization problem. I approach the task of finding globally optimal solutions to difficult nonconvex optimization problems by giving novel convex reparametrizations of nonconvex problems, using ideas from topology, geometry, and algebra. I will give examples of my past work exploiting such 'hidden convexity' in the context of manifold optimization, the design of algorithms, and other applications of 'log-concavity'.

## 1 Convex Projections of Manifolds

Optimization problems over complicated domains, such as nonconvex manifolds, are abundant in applications ranging from robotics and astronomy. However, in most applications, it is only necessary to understand the range of values of a small number of functions of the complicated variable: specifically those necessary to express the objective and constraints. From a geometric perspective, this means that the optimization problem can be re-expressed in terms of a low dimensional projection of the complicated domain; we are therefore interested in understanding when such projections give rise to simpler optimization problems.

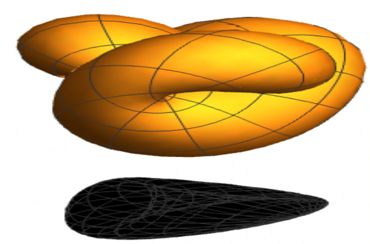


Figure 1: A depiction of a convex 2 dimensional shadow of a complicated nonconvex 3 dimensional shape.

As an example, in [10], I showed that any two dimensional linear projection of the set of rotation matrices, $\mathrm{SO}(n)$ is convex, a fact which has applications in astronomy for orientation finding.

In [6], I consider the following general question: *If $X$ is a set, and $f : X \to \mathbb{R}^k$ is a function, then under what conditions is the image $f(X)$ is convex?* If $f(X)$ is convex, then any optimization problem over $X$ which only depends linearly on $f$ can be formulated as a convex optimization problem. In order to tackle this question, I considered a Lagrangian associated to the function $f$: the function $\mathcal{L}(\lambda, x) = \langle \lambda, f(x) \rangle$. If $X$ were Euclidean space, and each $f_i$ were strictly convex, then it is clear that as the Lagrange multipliers $\lambda$ vary, the minimizer $x \in X$ of the function $\langle \lambda, f(x) \rangle$ changes continuously. This fact underlies a variety of algorithms such as interior point methods and augmented Lagrangians, which associate to a constrained optimization problem a family of unconstrained problems and track the solution of that family of problems as the multipliers vary. The main result of [6] is in some sense a converse to this fact:

**Theorem 1.1** (Informal)**.** *Let $X$ be a compact topological space and $f : X \to \mathbb{R}^k$ be continuous. If the minimizer of the Lagrangian, $\operatorname{argmin}_{x \in X} \langle \lambda, f(x) \rangle$, is continuous in $\lambda$, then $f(X)$ is convex.*

This result can easily recover many known results about hidden convexity, including Brickman's theorem, Dine's theorem, my earlier result on projections of the set of rotation matrices, as well as the main theorems of several papers on linear projections of the Stiefel manifold and quadratic maps on the sphere.

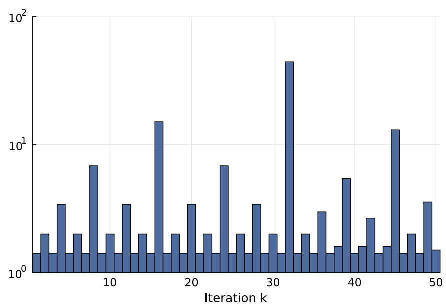
Algorithmically, I also showed that it is possible to solve problems with this property by tracking the continuous path of solutions to the Lagrangian optimization problem as the multipliers vary. I analyzed such an algorithm in the case where the domain $X$ is a Riemannian manifold.

## 2 Design of Algorithms via Convex Optimization

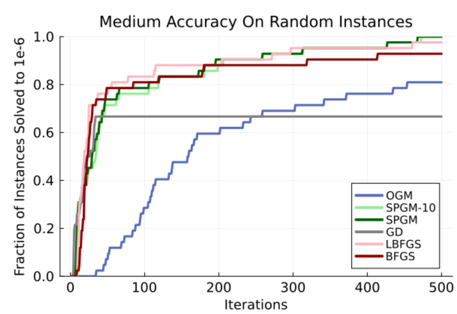
A first order method minimizes a function $f : \mathbb{R}^d \to \mathbb{R}$ using first order information about $f$, i.e. the value of $f$ and its gradients at a sequence of points $x_0, \ldots, x_n \in \mathbb{R}^d$. Such methods, including gradient descent, are cornerstones of modern optimization.

I am tackling the problem of *designing* first order methods for convex optimization, i.e. systematically understanding how to choose query points for optimal convergence rates.

In my first work on this subject, I considered query points defined by $x_{i+1} = x_i - h_i \nabla f(x_i)$ for fixed $h_0, \ldots, h_{n-1}$. Until recently, it was not known whether there exist step sizes $h_i$ which substantially improved the convergence rate of this method over constant step sizes. I showed in [8] that it is possible to improve on the guarantee offered by constant step size gradient descent by a polynomial factor using a sequence of unbounded and nonmonotonic step sizes. This has led to a flurry of activity on acceleration using improvements in step size choices, including our own follow up work showing that these step sizes are part of a much larger family of step sizes which have interesting algebraic properties[9].




(a) Step sizes for gradient descent with the best known convergence rate with 50 steps. Note that these step sizes are both nonmonotonic and not symmetric.




(b) Comparison of different algorithms against our method (SPGM) and limited memory versions (SPGM-10).

In [7], I introduced a new algorithm called SPGM which offer a beyond-worst-case guarantee for convex optimization. Specifically, I found an algorithm for minimizing a smooth convex function which has the property that after any number of iterations, the algorithm offers the best possible convergence guarantee *even amongst all algorithms which see those initial gradients and function values*. That is, before seeing any gradients, our algorithm offers a guarantee on its final suboptimality which is best possible, and as additional gradients come in, the algorithm is able to adjust to the new gradient information in an optimal fashion by solving a simple convex subproblem. Our results are also competitive with algorithmic staples such as BFGS in running time.

## 3 Applications of Log-Concavity

I have studied several basic linear algebra problems from the point of view of log-concave polynomials which have been the subject of much recent attention. I have studied such polynomials in their own right in [2, 3], but will focus on concrete applications here.

**Unbiased Regression using Random Eigenvalues** Suppose $\mu$ is a probability distribution on $\mathbb{R}$, and that we are given the ability to evaluate a function $f : \mathbb{R} \to \mathbb{R}$ at an arbitrary $x \in \mathbb{R}$. How can we find a polynomial $p$ of degree at most $d$ minimizing $\mathbb{E}_{x\sim\mu}[|p(x) - f(x)|^2]$, using as few samples points from $f$ as possible? We show in [5] that the following is true: take $X$ to be a $d+1\times d+1$ *random Hermitian* matrix whose density depends on $\mu$. Let $\lambda_0, \ldots, \lambda_d$ be the eigenvalues of $X$ (which are random). If we interpolate a polynomial $\hat{p}$ so that for each $i = 1, \ldots, d+1$, $\hat{p}(\lambda_i(X)) = f(\lambda_i(X))$, then $\mathbb{E}[\hat{p}]$ is the *optimal solution to the polynomial regression problem.* We also use random matrix theory to give both $O(d\log(d))$ time sampling algorithms and optimal sample complexity in many cases. Underlying this result is the theory of determinantal point processes, which are closely related to log-concave polynomials.

**Spot-checking PSDness** Suppose that we know that all $k \times k$ submatrices of a large $n \times n$ matrix $X$ are positive semidefinite (PSD). How far can $X$ be from being PSD? I showed in [1] that $\lambda_{min}(X) \geq -\frac{k-n}{n(k-1)}\mathrm{tr}(X)$, and that this is tight. Minimizing this objective over the space of matrices with PSD submatrices is nonconvex, but I was able to give a tight convex relaxation. This has applications in semidefinite programming, where rather than checking that a large matrix is PSD, it can be cheaper and more efficient to spot-check some small submatrices for being PSD. See also [4], where the same idea is used to understand sparse semidefinite programs.

## 4 Future Directions

My past work has demonstrated a number of incarnations of the idea of hidden convexity and its applications. My broad goal in the future is to develop a deeper systematic understanding of hidden convexity and to use that understanding to develop further applications.

One phenomenon that I am particularly interested in exploring is what I refer to as 'projection simplicity'. In convex optimization, there is a well established notion of 'extension complexity', in which a convex optimization problem (particularly a linear program) can be made simpler by actually lifting it into a higher dimensional space. I propose that we should also understand the 'dual' notion of projection simplicity, wherein a complicated high dimensional problem can be projected into a lower dimensional space which captures the relevant features of the problem but which yields a simpler optimization problem. My work on convex projections of manifolds is one example in which an intractable optimization problem can be projected into a simpler optimization problem, but I am also interested in broader questions along these lines. For example, rather than understanding whether or not a projection of a manifold is convex on the nose, I am interested in *approximate* notions of convexity, and whether the topological conditions I described can be made quantitative in such a way as to provide approximation results.

# References


[1] Grigoriy Blekherman, Santanu S Dey, Kevin Shu, and Shengding Sun. Hyperbolic relaxation of $k$-locally positive semidefinite matrices. *SIAM Journal on Optimization*, 32(2):470–490, 2022.

[2] Grigoriy Blekherman, Mario Kummer, Raman Sanyal, Kevin Shu, and Shengding Sun. Linear principal minor polynomials: hyperbolic determinantal inequalities and spectral containment. *International Mathematics Research Notices*, 2023(24):21346–21380, 2023.

[3] Grigoriy Blekherman, Julia Lindberg, and Kevin Shu. Symmetric hyperbolic polynomials. *Journal of Pure and Applied Algebra*, 229(2):107869, 2025.

[4] Grigoriy Blekherman and Kevin Shu. Sums of squares and sparse semidefinite programming. *SIAM Journal on Applied Algebra and Geometry*, 5(4):651–674, 2021.

[5] Chris Camaño, Raphael A. Meyer, and Kevin Shu. Debiasing polynomial and fourier regression. *arXiv preprint arXiv:2508.05920. To appear in SOSA 2026.*, 2025.

[6] Venkat Chandrasekaran, Tim Duff, Jose Israel Rodriguez, and Kevin Shu. Lagrangian dual sections : A new perspective on hidden convexity. *arXiv preprint arXiv:2510.06112*, 2025.

[7] Benjamin Grimmer, Kevin Shu, and Alex L Wang. Beyond minimax optimality: A subgame perfect gradient method. *arXiv preprint arXiv:2412.06731*, 2024.

[8] Benjamin Grimmer, Kevin Shu, and Alex L Wang. Accelerated objective gap and gradient norm convergence for gradient descent via long steps. *INFORMS Journal on Optimization*, 7(2):156–169, 2025.

[9] Benjamin Grimmer, Kevin Shu, and Alex L Wang. Composing optimized stepsize schedules for gradient descent. *To Appear in Mathematics of Operations Research*, 2025.

[10] Akshay Ramachandran, Kevin Shu, and Alex L Wang. Hidden convexity, optimization, and algorithms on rotation matrices. *Mathematics of Operations Research*, 50(2):1454–1477, 2025.