John M. Dedyo Yale University

The QCQP Approach, a Novel Design Strategy for Optical Computing Devices

I. INTRODUCTION

Artificial intelligence has a size problem. Existing infrastructure is struggling to keep up with the immense computational and energy demands of ChatGPT-4 and other cutting-edge AI models. Optics and photonics has the potential to provide a solution.

For AI computations, current digital computers can require eight of the highest performance chips available (GPUs) which, alongside other computational requirements, often warrant the use of a dedicated data center. Microsoft has even sought to restart the Three Mile Island nuclear plant to offset the huge demands its AI facilities impose on the electrical grid. [1] By comparison, an optical computer small enough to fit on a desk could implement machine learning models the size of ChatGPT-4 while consuming orders-of-magnitude less energy. [2] However, existing approaches to the computational design of such electromagnetic wave-based optical computing systems require impractically high computing power and often produce suboptimal designs. Thus, these devices are currently fashioned manually, resulting in prohibitively large device footprints. Such a manually designed optical device implementing an operation about 1% of the size of a typical image would require a footprint about 50 *meters* long [3], a scale that is clearly infeasible.

In this project, I worked independently to implement and evaluate a novel photonic optimization paradigm for the computational design of highly compact optical devices. The approach recasts the typical photonic design problem—an optimization problem constrained by Maxwell's equations—as a quadratically-constrained quadratic program (QCQP), thereby addressing the computational inefficiency standing in the way of traditional gradient-based approaches to designing compact devices.

II. METHODOLOGY

As the sole researcher, I devised and studied a binary-choice optimization problem that seeks to maximize the power of an electromagnetic wave measured at a single point x^* by organizing a grid of electric dipoles (Eq. (1)).

$$\begin{aligned} \underset{\varrho}{\text{maximize}} \quad & f(\varrho) = |E_{\text{net}}(x^*; p(\varrho))|^2 \\ \text{subject to} \quad & \varrho_i \in \{\varrho_a,\ \varrho_b\} \text{ for } i = 1, 2, \ldots, \\ & (\mathbb{I} - \alpha_0 \text{diag}(\varrho)\mathbb{G})p = \alpha_0 \text{diag}(\varrho) E_{\text{inc}} \end{aligned} \tag{1}$$

The first constraint requires each dipole to be in one of two states ϱ, either ϱ_a or ϱ_b. This is the "binary-choice" variable. The waves emitted by the dipoles in response to the driving electromagnetic wave, E_{inc}, are permitted to interact with each other and with the other dipoles, governed by the second constraint. These interactions determine the polarization p of each dipole (which is calculated by inverting the second constraint and thus a function of ϱ). The coefficient α_0 is simply the linear polarizability of the dipoles (i.e., what material they are), and $\mathbb{G}$ is a matrix of the dipoles' Green's functions.

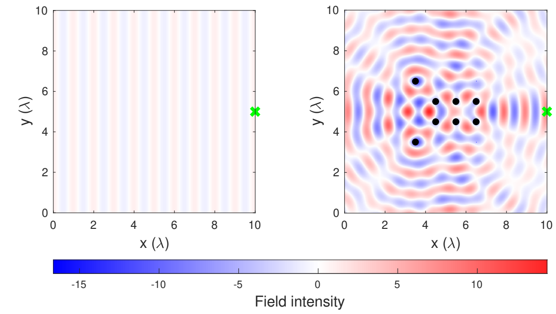

FIG. 1. Scattering system before (left) and after (right) optimization with the QCQP approach. On the left, the relatively weak driving field E_{inc} is parallel to the x axis and has magnitude 1. After optimization, the dipoles (black dots) emit their own scattered waves, each driven by E_{inc} and the other dipoles' scattered waves. The optimized grid is laid out to create constructive interference in E_{net} focused at the target (green "X").

Figure 1 shows the system before (left) and after (right) solving Eq. (1) using the QCQP approach described in Section II C. Power at the green "X" is maximized by adding dipoles (black circles). The optimized "degree of freedom" is whether each dipole is present ($\varrho_i = 1$) or absent ($\varrho_i = 0$). This problem has a similar degree of difficulty and nonconvexity as many others in wave optics, so it is a strong testbed for techniques that can be used to design true optical computing devices.

A. Traditional Gradient-Based Optimization

To start, I constructed a MATLAB simulation that uses the dyadic Green's function in 2D to calculate the electric field, E_{net}, generated by the dipoles in response to a driving electromagnetic wave, E_{inc}. Then using linear algebra techniques and the $\mathbb{CR}$-calculus [4], I derived the adjoint equations and accompanying p_{adj} to formulate a computationally efficient optimization approach requiring only one additional simulation to calculate the gradient (Eq. (2)) rather than the thousands of additional simulations required for numerical approaches. To solve the optimization problem, I used a standard gradient descent routine, MATLAB's `fmincon` function with the interior-point algorithm.

$$\frac{\partial f}{\partial \varrho_i} = 2\text{Re}\left[\alpha_0 \left(E_{\text{inc}} + \mathbb{G}p\right)_i p_{\text{adj},\, i}\right] \tag{2}$$

This technique, the adjoint state method, optimizes by gradient descent, which is often caught in low-quality *local* optima. Currently the dominant optimization paradigm in photonic design, this method's performance provides an important point of comparison for the novel QCQP approach.

B. SDP Optimization for Fundamental Bounds

I used linear algebra techniques to rederive this optimization problem (Eq. (1)) as a QCQP, which is a well-studied class of computational problems that have not before been applied to photonic design. [5] The QCQP approach formulates the problem in terms of a matrix variable, $P = pp^\dagger$. The operators $\mathbb{A}_0$ and $\mathbb{B}_i$ encode the objective function and constraints, respectively, from Eq. (1). This reformulation concentrates the difficulty (that is, nonconvexity) of the problem into a single optimization constraint: that the matrix, P, has rank one. In a strategy called a "semidefinite relaxation," I dropped the difficult constraint while maintaining all others (Eq. (3)), and then computationally solved this easier problem, called a semidefinite program (SDP), using the `cvx` package in MATLAB. The semidefinite relaxation is closely related to the original problem and provides a fundamental limit on its optimal value, but achieving this optimal value in a real device is often a physical impossibility.

$$\begin{aligned} \underset{P}{\text{maximize}} \quad & \operatorname{Tr}(\mathbb{A}_0 P) \\ \text{subject to} \quad & \operatorname{Tr}(\mathbb{B}_i P) = 0 \quad \text{for } i = 1, 2, \ldots, \\ & P \geq 0 \end{aligned} \tag{3}$$

C. Physically-feasible QCQP Optimization

Of course, the difficult constraint ultimately is important to the physical reality of the optimization problem, so it must be accounted for when looking for a design that can actually be built in the physical world. To solve this more difficult problem, I implemented the "majorization-minimization" method [6] that iteratively solves a closely related SDP with the same constraints as the SDP in Eq. (3), but with a slightly altered (i.e., rank-penalized) objective function. At each step, the rank penalty is increased, so the solution to this rank-penalized SDP approaches a solution to the original, physically-realistic QCQP including the difficult rank-one matrix constraint (Eq. (4)).

$$\begin{aligned} \underset{P}{\text{maximize}} \quad & \operatorname{Tr}(\mathbb{A}_0 P) \\ \text{subject to} \quad & \operatorname{Tr}(\mathbb{B}_i P) = 0 \quad \text{for } i = 1, 2, \ldots, \\ & P \geq 0, \\ & \operatorname{rank} P = 1 \end{aligned} \tag{4}$$

III. RESULTS

Recasting this computationally difficult (nonconvex) photonic design problem as a quadratically-constrained quadratic program appears to lead to improved optimization performance when compared to the gradient-descent-based adjoint state method, the dominant paradigm in photonics.

Numerical experimentation shows that the QCQP approach reliably finds higher quality optima than does gradient descent, even when gradient descent is run 100 or more times from different random starting points. Figure 2 compares the optimization performance of the adjoint state method (gradient descent), the SDP, and the QCQP approaches in 100 trials on the same problem each seeded with a random starting point. As this is a maximization problem, higher objective values correspond to better performance.

The "optimization landscape" for this wave-based problem is highly oscillatory, so the gradient descent optimization routine is unreliable, which is evident in the oscillations of the red line (Fig. 2). That is, the optimization performance varies widely between runs when the optimization is seeded with a random starting point. Ideally in optimization, one can be confident that at least a near-global optimum has been reached whenever the routine terminates, but gradient descent does not offer such assurances here. It is also clear that the SDP solution (green line) is a fundamental limit on performance as it is never exceeded, and that the QCQP result is generally superior to that of gradient descent. The SDP and QCQP lines are horizontal since the performance of both techniques is independent of the starting point, unlike gradient descent.

As problem complexity increases, the QCQP approach continues generally to outperform the gradient-based adjoint state method. Fig. 3 shows the optimization objective value as a function of the number of dipoles in the grid being optimized (a proxy for problem difficulty), and the error bars on the red line give a two standard deviation confidence interval from 100 randomly-seeded gradient descent optimizations for the same problem. As in Fig. 2, higher values indicate superior performance. The lines all meet at 1 dipole, which is to be expected because there is little to optimize in such a problem—each method naturally converges to an optimal solution that includes the standalone dipole. The QCQP approach (blue line) outperforms gradient descent (red line) more dramatically in more difficult problems, which indicates

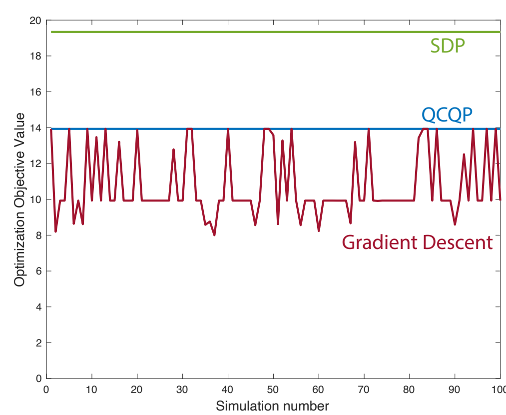

FIG. 2. Optimization performance for the gradient descent (adjoint state), QCQP, and SDP problems from 100 randomly-seeded start points (i.e., initial dipole layouts). The SDP approach provides a fundamental bound on the optimal objective value (y axis). The QCQP approach reliably converges to a high-quality and physically-feasible optimum. The gradient descent method is unreliable, oscillating between high- and low-quality optima.

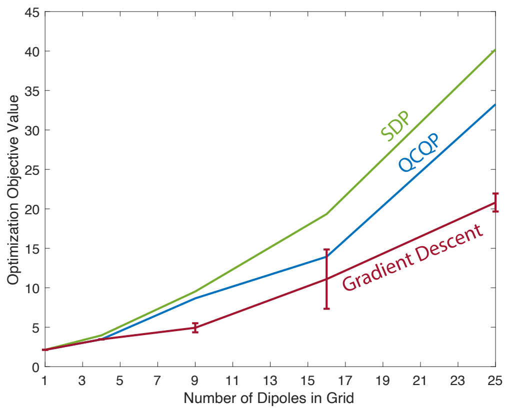

FIG. 3. Optimization performance for the gradient descent (adjoint state), QCQP, and SDP approaches on problems with increasing difficulty (i.e., more dipoles to position in the grid). The gradient descent optima are, on average, worse than those of the QCQP, and the error bars give a two standard deviation confidence interval from 100 randomly-seeded gradient descent optimizations on the same problem. The SDP and QCQP lines diverge more rapidly from the gradient descent line as the problem becomes more complex, reflecting gradient descent's inability to cope with the most difficult problems.

its promise as a paradigm for designing highly sophisticated wave scattering devices like optical computers. Indeed, it is the poor performance of gradient descent in these difficult problems that has necessitated the manual design of optical computing devices.

While not a conclusive mathematical proof, these results indicate that recasting a difficult photonic design problem as a QCQP improves optimization performance and could play a role in a method of computationally designing compact yet fully-sophisticated optical computers to meet AI's exploding computational demands.

IV. CONCLUSION AND FUTURE DIRECTIONS

The great advantage of optical computing devices is that they leverage the properties of wave scattering to perform matrix-vector products entirely in the physical domain, making them highly energy- and time-efficient. But this wave behavior makes typical methods of computational design—those based on gradient descent—poorly-suited to solve this wave-optics problem because the optimization landscape becomes highly oscillatory as the design complexity increases (that is to say, there are lots of "peaks and valleys" that will cause the gradient descent optimization routine to terminate at a point that is not globally optimal; see Fig. 2). Yet human-intuition based designs are prohibitively large, and better-behaved gradient-based alternatives to gradient descent are too computationally intensive for designing the most complex optical devices. The QCQP approach that I use here, on the other hand, "smoothes out" the optimization landscape, resulting in more reliable convergence to a high-quality solution.

These findings contribute to a larger vision of a novel design methodology that computationally designs reasonably-sized optical computing devices by (1) decomposing the device into simple, uncoupled components and then (2) designing each component computationally. Although the QCQP approach is currently incapable of designing the most complicated devices in a single shot, it has strong potential as a method of designing this series of separate, uncoupled sub-devices. Such a technique could efficiently produce highly compact, yet fully sophisticated optical computing devices with substantial advantages over digital devices.

A number of future inquiries are required to realize the ultimate project aim of creating a fully-optical AI device. My next step will be to endow the QCQP approach with an optimization objective that designs a bona fide optical computing device, as well as a more physically realistic representation of a dipole (including the magnetic dipole moment and precise measures of material lossiness). Then to design more complex devices, I will devise an algorithmic strategy for breaking down a large device into simple, uncoupled sub-devices, each of which can then be designed independently with the QCQP approach and assembled into a compact yet highly sophisticated optical computer, as noted above. With the computational design methodology complete, I will collaborate with an experimental research group to manufacture these computationally designed sub-devices and assemble them into an actual compact optical computer. After that, we will be ready to integrate optical devices into the existing AI ecosystem as an environmentally-friendly alternative to large data centers.

[1] M. Greshko, Nuclear power for ai: what it will take to reopen three mile island safely, Nature **634**, 272 (2024).

[2] E. Cartlidge, Scattered light holds promise for efficient ai (2024).

[3] H. Larocque and D. Englund, Universal linear optics by programmable multimode interference, Opt. Express **29**, 38257 (2021).

[4] K. Kreutz-Delgado, The complex gradient operator and the cr-calculus (2009), arXiv:0906.4835 [math.OC].

[5] S. Gertler, Z. Kuang, C. Christie, and O. D. Miller, Many physical design problems are sparse qcqps (2023), arXiv:2303.17691 [physics.optics].

[6] T. Liu, B. Sun, and D. H. K. Tsang, Rank-one solutions for sdp relaxation of qcqps in power systems, IEEE Transactions on Smart Grid **10**, 5 (2019).