Found Comput Math
DOI 10.1007/s10208-014-9197-9

FOUNDATIONS OF COMPUTATIONAL MATHEMATICS
The Journal of the Society for the Foundations of Computational Mathematics

Extensions of Gauss Quadrature Via Linear Programming

Ernest K. Ryu · Stephen P. Boyd

Received: 15 May 2013 / Revised: 18 January 2014 / Accepted: 10 February 2014

Abstract Gauss quadrature is a well-known method for estimating the integral of a continuous function with respect to a given measure as a weighted sum of the function evaluated at a set of node points. Gauss quadrature is traditionally developed using orthogonal polynomials. We show that Gauss quadrature can also be obtained as the solution to an infinite-dimensional linear program (LP): minimize the nth moment among all nonnegative measures that match the 0 through $n-1$ moments of the given measure. While this infinite-dimensional LP provides no computational advantage in the traditional setting of integration on the real line, it can be used to construct Gauss-like quadratures in more general settings, including arbitrary domains in multiple dimensions.

Keywords Gauss quadrature · Semi-infinite programming · Convex optimization

Mathematics Subject Classification 65D32 · 90C34 · 90C48

Communicated by Michael Overton.

E. Ryu (✉)
Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305, USA
e-mail: eryu@stanford.edu

S. Boyd
Electrical Engineering, Stanford University, Stanford, CA 94305, USA
e-mail: boyd@stanford.edu

Published online: 25 March 2014

Springer

1 Gauss Quadrature

We briefly review Gauss quadrature and set up our notation. Let $\Omega \subset \mathbf{R}$ be a closed interval and q a given measure on Ω. The standard method for approximating the definite integral of a continuous function f on Ω is

$$\int_\Omega f(x)\, dq(x) \approx \sum_{j=1}^{N} w_j f(x_j).$$

The right-hand side is referred to as a *quadrature*. The coefficients $w_1, w_2, \ldots, w_N$ are the *weights* and $x_1, x_2, \ldots, x_N \in \Omega$ are the *nodes*, i.e., the locations at which the function f is sampled to form the approximation. The quadrature is said to be of *order* n if it is exact for polynomials up to degree $n-1$, i.e.,

$$\int_\Omega x^i\, dq = \sum_{j=1}^{N} w_j x_j^i, \quad i = 0, \ldots, n-1.$$

The numbers on the left-hand side are the 0 through $n-1$ moments of the measure dq. These conditions are a set of n linear equations in the N weights. For $N = n$ (and for $N \geq n$), for any choice of distinct nodes, we can always find weights that satisfy the preceding equations since the coefficient matrix for the linear equations is Vandermonde and, therefore, invertible. (However, the resulting weights are not necessarily nonnegative.) Thus, a quadrature of order n can be found by choosing an arbitrary set of distinct $N = n$ nodes. We call a quadrature of order n with $N < n$ nodes *efficient*; such a quadrature requires fewer function evaluations than its order. The linear equations for the weights of an efficient quadrature have more equations than variables; these equations are not solvable unless the nodes are chosen very carefully.

In 1814 Gauss [11] discovered the first efficient quadrature, which is now called a *Gauss quadrature*. A Gauss quadrature of order n requires only $N = n/2$ nodes (for n even). Traditionally a Gauss quadrature is developed with the theory of orthogonal polynomials; such a treatment can be found in many standard texts [19,9,32]. There are efficient methods to find Gauss quadrature nodes and weights, such as the Golub–Welsch algorithm [14] and the Glaser–Liu–Rokhlin algorithm [12].

2 Gauss Quadrature Via Linear Programming

As we show in this paper, a Gauss quadrature can also be obtained as the solution of an infinite-dimensional linear program (LP) over nonnegative measures. Again, let $\Omega \subseteq \mathbb{R}$ be a closed (but not necessarily compact) interval. Assume that $\mathbf{supp}\, q = \Omega$, where $q \geq 0$ is the given nonnegative measure of integration to approximate, and that n is even, and consider the optimization problem

$$\begin{aligned} &\text{minimize} \quad \int_\Omega x^n \, d\mu \\ &\text{subject to} \quad \int_\Omega x^i \, d\mu = \int_\Omega x^i \, dq, \quad i = 0, \ldots, n-1, \mu \geq 0, \end{aligned} \tag{1}$$

where $\mu \in \mathcal{M}$ is the optimization variable and $\mathcal{M}$ the space of finite Borel measures on Ω. This is a LP with n equality constraints and an infinite-dimensional variable, the measure μ. Problem (1) seeks a nonnegative measure with smallest nth moment while matching the 0 to $n-1$ moments of dq.

Theorem 1 *There is a unique solution $\mu^\star$ to the LP (1) given by*

$$\mu^\star = \sum_{i=1}^{n/2} w_i \delta_{x_i},$$

where $w_1, \ldots, w_{n/2}$ and $x_1, \ldots, x_{n/2}$ are the weights and nodes of the Gauss quadrature and δ_{x_i} denotes the Dirac measure.

By analogy to basic feasible solutions of finite-dimensional linear programs [22, §2.4], one may expect $\mu^\star$ to be discrete with $|\mathbf{supp}\, \mu^\star| \leq n$. Moreover, since the constraint $\mu \geq 0$ enforces $w_i \geq 0$ for all i, it is not surprising that $\mu^\star$ is a quadrature with positive weights. What is surprising is that $\mu^\star$ is in fact a Gauss quadrature (which has $|\mathbf{supp}\, \mu^\star| = n/2$). The proof of Theorem 1 is given in the appendix.

We immediately point out that this observation gives no computational advantage at all in the univariate setting; it is certainly simpler to compute Gauss quadrature nodes and weights using the classical methods than by solving an infinite-dimensional LP over the space of nonnegative measures. The advantage of the LP formulation is that it generalizes to other settings, as we will explore in §3.

We can give $\mu^\star$ a minimum sensitivity interpretation. Consider a polynomial of degree n, $f(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n$. Then for any feasible μ,

$$\int_\Omega f(x) \, d\mu = \alpha_0 \int_\Omega 1 \, d\mu + \alpha_1 \int_\Omega x \, d\mu + \cdots + \alpha_n \int_\Omega x^n \, d\mu$$

holds, and the objective in (1) gives the sensitivity of the quadrature to α_n. Thus the LP can be interpreted as seeking the measure that gives the exact integral for polynomials of degree less than n and is least sensitive to the x^n term.

We can also interpret the optimization problem (1) as a (weighted) ℓ_1-norm minimization problem. Adding a constant α to the integrand x^n ensures that the integrand is strictly positive, without changing the problem (since the integral of a constant is fixed by the moment constraint), and we can do the same for odd n if the domain is bounded. The objective can then be written as $\int_\Omega (\alpha + x^n) \, d|\mu|$ since μ is nonnegative. Minimizing a (possibly weighted) ℓ_1-norm to obtain a sparse solution (in this case,

one with finite and small support) is the central idea of compressed sensing [6] and many other related methods such as lasso [33] and basis pursuit [7].

We conclude this section with two quick remarks. First, if in (1) we maximize instead of minimize, then we obtain a Lobatto quadrature. If n is odd instead of even, then we obtain a Randau quadrature [1, p.888]. Second, Theorem 1 can be generalized to Chebyshev systems, a set of equations that are in some sense like polynomials [17,28,24,18]. We omit the proofs as they are straightforward modifications of the main result.

3 Extensions of Gauss Quadrature Via Linear Programming

We observe that the LP approach makes sense in a more general setting. Let $\Omega \subset \mathbf{R}^d$ be a compact domain, with $C(\Omega)$ denoting the space of continuous functions on Ω. In analogy to the powers $x^0, \ldots, x^{n-1}$ that appear in (1), we let $p^{(0)}, \ldots, p^{(n-1)} \in C(\Omega)$ be a linearly independent set of *test functions*, with $p^{(0)} = 1$. We let $r \in C(\Omega)$ be a function that will serve the role of x^n in the LP (1); we refer to it as the *sensitivity function* and assume it is linearly independent of the test functions.

Let q be a nonnegative Borel measure with $\mathbf{supp}\, q = \Omega$. For convenience, we use the notation $\mu(f) = \int_\Omega f \, d\mu$ for any $f \in C(\Omega)$ and $\mu \in \mathcal{M}$.

We seek quadratures that approximate q, i.e.,

$$q(f) \approx \sum_{j=1}^{N} w_j f(x_j),$$

where $x_1, \ldots, x_N$ are the nodes and $w_1, \ldots, w_N$ the weights. We say a quadrature is of order n if it is exact on the test functions $p^{(0)}, \ldots, p^{(n-1)}$, i.e.,

$$q(p^{(i)}) = \sum_{j=1}^{N} w_j p^{(i)}(x_j), \quad i = 0, \ldots, n-1.$$

The motivation is similar to that of a Gauss quadrature; such a quadrature is accurate for functions that are approximated well by the test functions. As with standard quadratures on an interval, given a set of nodes $x_1, \ldots, x_N$, the aforementioned constraints are a set of n linear equations in the N weights; they are generically solvable for $N \geq n$, over choices of distinct nodes $x_1, \ldots, x_N$. We say a quadrature is efficient when $N < n$; in this case, the linear equations for the weights have more equations than unknowns, and so have no solution, except when the nodes are chosen carefully.

Motivated by the LP (1) we form the LP

$$\begin{array}{ll} \text{minimize} & \mu(r) \\ \text{subject to} & \mu(p^{(i)}) = q(p^{(i)}), \quad i = 0, \ldots, n-1, \\ & \mu \geq 0, \end{array} \tag{2}$$

where $\mu \in \mathcal{M}$ is the optimization variable. Here we seek a nonnegative measure that matches the values of the given measure on the test functions and, among all such measures, has a minimum value on the sensitivity function.

Theorem 2 *The LP* (2) *has a solution* $\mu^\star$ *that satisfies* $|\mathbf{supp}\,\mu^\star| \leq n$.

This theorem tells us that there is a solution with support on no more than n points; such a measure gives a quadrature with nonnegative weights on no more than $N = n$ nodes and with nodes within the domain Ω. Again, $\mu^\star$ is like a basic feasible solution of a finite-dimensional LP [22, §2.4]. The bound $|\mathbf{supp}\,\mu^\star| \leq n$ is tight in certain cases, so we cannot say more (for example, that there exists an efficient quadrature) without adding more assumptions about the given measure, the domain, the test functions, and the sensitivity function. But in many examples, the LP (2) produces efficient quadratures, analogous to a Gauss quadrature. We also note that the existence of quadratures of order n is clear, indeed, for a generic choice of nodes; the theorem says that there is a choice of no more than n nodes that yields a quadrature of minimum sensitivity.

The LP (1) that characterizes Gauss quadratures has a unique solution, but the generalized LP (2) can have multiple solutions. Moreover, it can have solutions with infinite support; we will see an example in §5.5. It is only when $\mathbf{supp}\,\mu^\star$ is a finite set that we can identify it with a quadrature, and the quadrature is efficient only if $|\mathbf{supp}\,\mu^\star| < n$.

We will refer to a quadrature obtained from the LP (2) as a *Gauss-LP quadrature*. Unlike a standard Gauss quadrature, such quadratures need not be unique; there can be multiple Gauss-LP quadratures for a given Ω, q, $p^{(0)}, \ldots, p^{(n-1)}$, and r.

Finally, we shall note that the choice of r is somewhat arbitrary. As discussed briefly in the conclusion, different choices of r yield different quadratures. This in particular tells us that in problem (2) we can either maximize or minimize because minimizing $\mu(-r)$ is equivalent to maximizing $\mu(r)$.

4 Numerical Methods

We first point out that the optimization problem (2) is convex, but in general NP-hard, when the dimension d is allowed to vary. The problem of deciding polynomial nonnegativity in $\mathbf{R}^d$ is NP-hard [25], and we will reduce it to problem (2) with Lasserre's approach to convexifying the polynomial nonnegativity problem [20].

Let r be a multivariate polynomial, $n = 1$, and $p^{(0)} = 1$. Then problem (2) becomes

$$\begin{aligned} &\text{minimize } \mathbf{P}(r) \\ &\text{subject to } \mathbf{P} \text{ is a probability measure.} \end{aligned}$$

The optimal $\mathbf{P}$ is supported on the points that minimize the polynomial, and the optimal value is the minimum value of r. This minimum value is nonnegative if and only if $r(x) \geq 0$ for all $x \in \mathbf{R}^d$. Therefore, we have reduced an NP-hard problem to problem (2); this implies that problem (2) is NP-hard and that there is no known efficient algorithm to solve it.

On the other hand, our interest is limited to cases with d fixed and quite small, say, 2 or 3, in which case there are effective methods for solving (2). Several standard methods can solve such infinite-dimensional optimization problems when d is small. One approach focuses on the dual problem, which has a finite number of variables but an infinite number of constraints and so is called a *semi-infinite program* [4,10]. A cutting-plane method can be used to solve the dual, from which we can construct a solution of the original (primal) problem. There are also algorithms that resemble the simplex or exchange method that directly solve the original problem [3,13].

For the sake of completeness we describe a simple but effective method for solving (2) when d is small, say, 2 or 3. Our description is informal; for formal descriptions of an algorithm to solve the infinite-dimesional LP we refer the reader to the references cited earlier.

We choose a finite set of sample points $\mathcal{S} = \{s_1, \ldots, s_M\} \subset \Omega$ (chosen to form a grid with small mesh size in Ω) and restrict μ to the finite-dimensional subspace of measures that are supported on $\mathcal{S}$ to obtain the problem

$$\begin{array}{ll} \text{minimize} & \mu(r) \\ \text{subject to} & \mu(p^{(i)}) = q(p^{(i)}), \quad i = 0, \ldots, n-1, \\ & \mu \geq 0, \\ & \mathbf{supp}\, \mu \subseteq \mathcal{S}, \end{array} \tag{3}$$

with variable $\mu \in \mathcal{M}$. If we represent μ using $\mu = \sum_{i=1}^{M} \alpha_i \delta_{s_i}$, this problem reduces to an ordinary finite-dimensional LP for the (nonnegative) variables $\alpha_1, \ldots, \alpha_M$, which is readily solved. The solution gives a quadrature using nodes contained in the sample set $\mathcal{S}$. Any basic feasible solution of the LP (say, the solution found using the simplex algorithm) has at most n nonzero coefficients. This gives us an order n quadrature, with at most $N = n$ nodes.

What we observe is that the support of the discretized LP (3) often contains $N < n$ clusters of sample points, near each point in the support of an optimal measure. We identify these N clusters, and for each cluster we choose a node point given by the weighted convex combination (in which the weight of x_i is proportional to w_i) of the nodes of the approximate quadrature within the cluster; we choose as the weight the sum of the weights in the cluster. We now have an approximate but efficient quadrature, with nodes $\hat{x}_1, \ldots, \hat{x}_N$ and weights $\hat{w}_1, \ldots, \hat{w}_N$.

To further refine our solution, we now switch to local optimization and solve the nonlinear least-squares problem

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{i=0}^{n-1} \left(q(p^{(i)}) - \sum_{j=1}^{N} w_j p^{(i)}(x_j) \right)^2 \\ \text{subject to} & x_i \in \Omega, \quad i = 1, \ldots, N, \\ & w_i \geq 0, \quad i = 1, \ldots, N, \end{array}$$

with variables $w_1, \ldots, w_N$ and $x_1, \ldots, x_N$, starting from our approximate solution $\hat{w}_1, \ldots, \hat{w}_N$ and $\hat{x}_1, \ldots, \hat{x}_N$. Using standard sequential quadratic programming with

an active set method [26], this typically converges quickly to a point with objective zero, and when it does (and if $N < n$), we have an efficient quadrature.

Finally, we note that while the method sketched here sounds heuristic (especially the step in which we identify N clusters), we can certify the final solution obtained as being optimal for (2) using its dual (5), given in the appendix. The certification requires that we check that a linear combination of $p^{(i)}$ and r is nonnegative on Ω, which can be done by fine sampling.

5 Examples

When our method is applied to integration on the real line, with polynomial test functions, a classical Gauss quadrature is recovered exactly, as predicted by Theorem 1. In the remainder of this section we report numerical results for our quadrature construction method on some more interesting examples in $\mathbf{R}^2$ and $\mathbf{R}^3$.

Traditional Gauss quadrature does not easily generalize to multidimensional integrals, and while much effort has been dedicated to this problem, the theory is far from complete. In particular, all known methods do not have optimality guarantees, although many perform very well in practice [27,15,30,31,8,35,34]. The purpose of this section is to provide a proof of concept for our method applied to this multidimensional setting. The work should not be considered an exhaustive investigation of the method's practical performance.

5.1 Gauss Quadrature on the Unit Disk in $\mathbf{R}^2$

We take $\Omega = \left\{(x, y) \in \mathbf{R}^2 : x^2 + y^2 \leq 1\right\}$, with measure $q(f) = \int_\Omega f \, dxdy$. We use polynomial test functions $x^p y^q$ for $p+q < m$, so $n = m(m+1)/2$, and the sensitivity function $r = x^m + y^m$. It is also possible to include other degree m monomials in r, and doing so would give us a different quadrature.

The resulting Gauss-LP quadrature is shown in Fig. 1 for the case $m = 10, n = 55$. The previously described method finds a solution of the LP (2) with support size 21, i.e., a quadrature with $N = 21$ nodes.

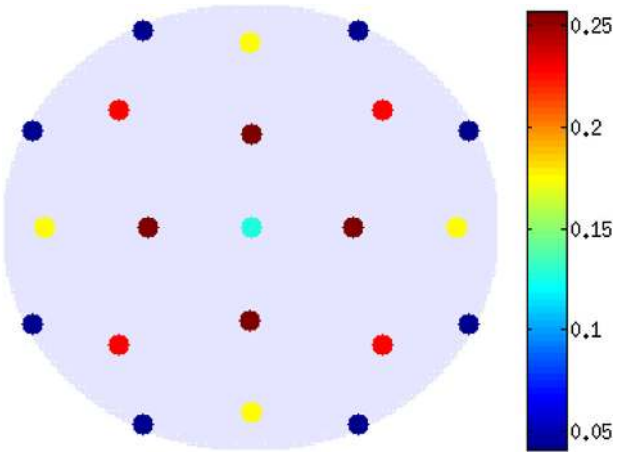


Fig. 1 Gauss-LP quadrature on unit *circle* in $\mathbf{R}^2$ with $m = 10$ and $n = 55$, with $N = 21$ nodes

Table 1 Number of nodes in Gauss-LP quadratures and Pierce's [27] quadratures

n	3	10	21	36	55	78	105	136	171	210
Gauss-LP	1	4	9	20	21	36	37	57	65	80
Pierce		4		16		36		64		100

Pierce's quadratures are only defined for every other entry

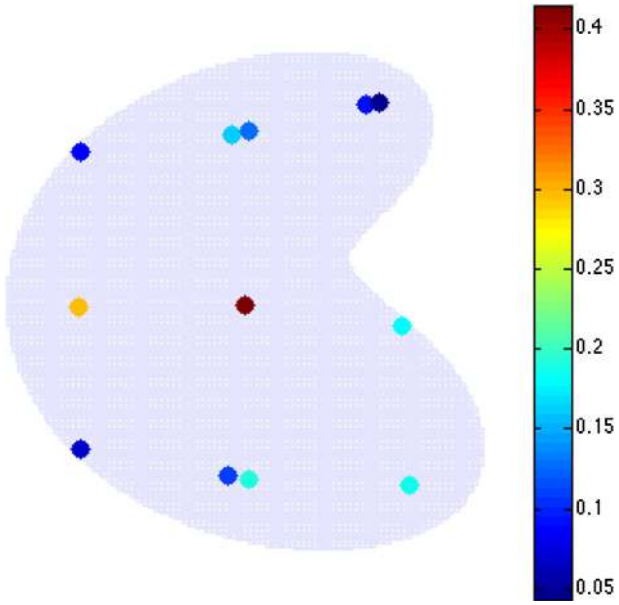


Fig. 2 Gauss-LP quadrature for $m = 6$ and $n = 21$, with $N = 12$ nodes

The Gauss quadrature for the unit disk is well studied; Pierce [27] gives a formula for quadratures for $\mathbf{R}^2$; more general formulas for $\mathbf{R}^d$ can be found in [31]. These quadratures rely on the product Gauss quadrature and the polar coordinate parameterization that maps $[-1, 1] \times [-\frac{\pi}{2}, \frac{\pi}{2})^{d-1}$ onto the unit ball. In Table 1 we compare the number of required nodes, given the same test functions, between Pierce's quadratures and Gauss-LP quadratures found using the method described earlier. It appears that the Gauss-LP quadratures are at least competitive with, and for larger orders more efficient than, Pierce's quadratures.

5.2 Gauss Quadrature on Arbitrary Domain in $\mathbf{R}^2$

In this example (and the next) we look at a quadrature on a nonconventional domain, with Ω defined via a polynomial inequality,

$$\Omega = \left\{ (x, y) \in \mathbf{R}^2 : \left(x^2 + y^2\right)^2 + (1/2)y^3 \leq x\left(x^2 + 4y^2\right) \right\}. \tag{4}$$

For our first example, the given measure is simple integration on Ω, $q(f) = \int_\Omega f \; dxdy$. The test functions are monomials of degree less than m, and the sensitivity function is $r = x^m + y^m$.

The Gauss-LP quadrature found for the case $m = 6$, $n = 21$ is shown in Fig. 2. It has $N = 12$ nodes, a bit more than half the order.

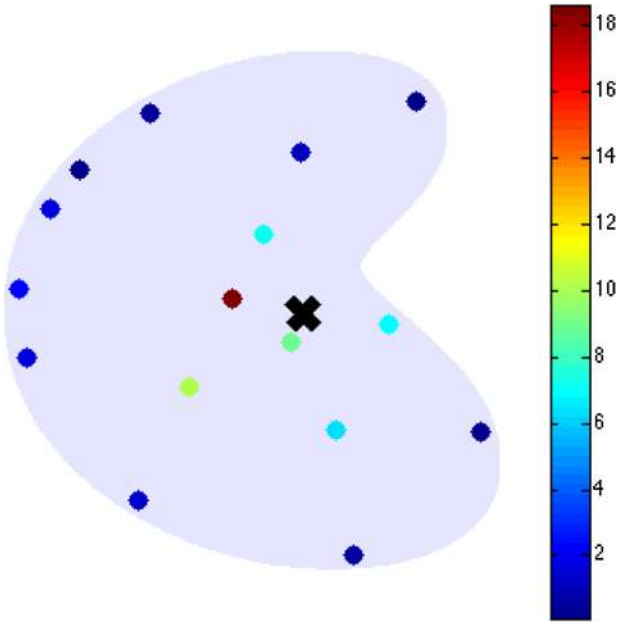


Fig. 3 Gauss-LP quadratures for Poisson equation with $n = 30$ and 16 nodes. *Black cross*: location of x_0

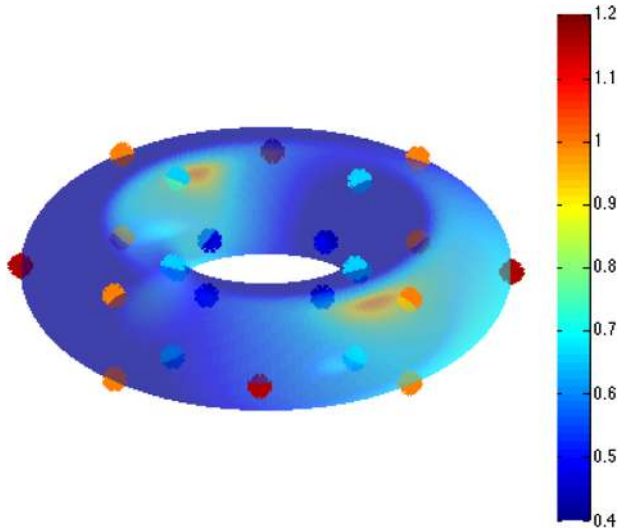


Fig. 4 Gauss-LP quadrature for $m = 6$ and $n = 56$, with $N = 24$ nodes

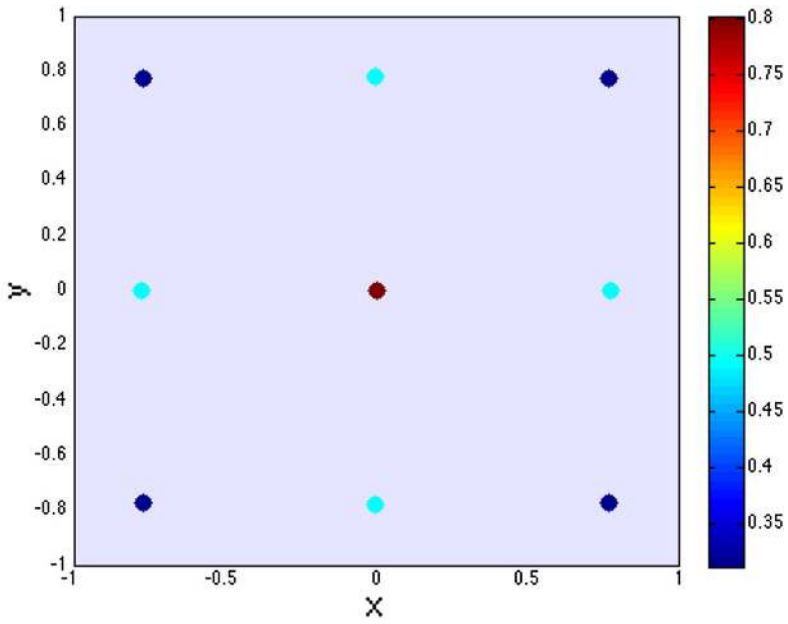
In Fig. 2 we see several paired nodes, which raises the question of whether they are numerical artifacts. However, they are not; the computed measure is indeed the solution to problem (2), and this is verifiable, as discussed in Sect. 4.

It may be possible to improve the quadrature by further reducing the number of nodes, perhaps by joining the pairs and running Newton's method for root finding [35,23]. Combining our method with existing ones could be an interesting direction for future work.

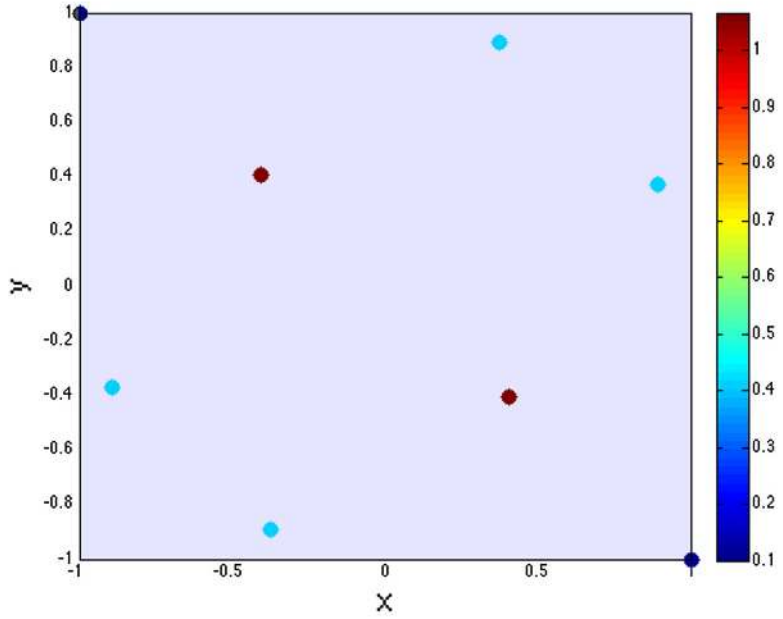
5.3 Gauss-LP Quadrature to Approximate the Poisson Equation

Let $\phi_f : \Omega \to \mathbf{R}$ be the solution to the Poisson equation

$$-\nabla^2 \phi_f = f \text{ on } \mathbf{int}\,\Omega, \qquad \phi_f = 0 \text{ on } \partial\Omega,$$



(a) Product Gauss quadrature obtained with $r = x^m + y^m$.



(b) More efficient quadrature obtained with different r.

Fig. 5 Gauss-LP quadrature on 2D unit square with $m = 6$ and $n = 21$. The first quadrature uses 9 nodes, whereas the Gauss-LP quadrature produced with the rather odd choice of r uses only 8

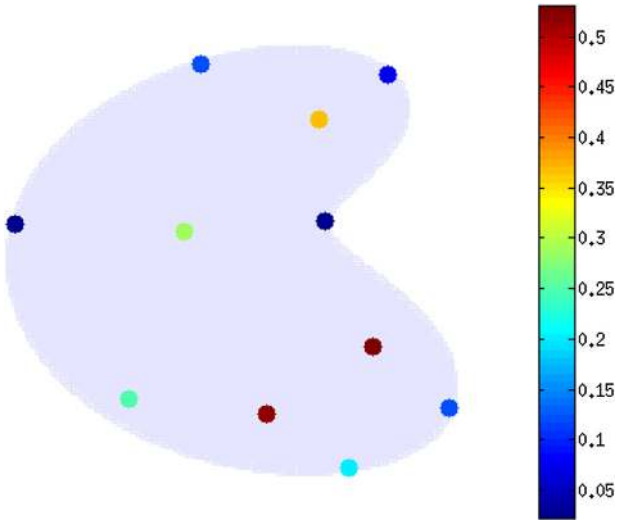


Fig. 6 More efficient quadrature obtained with different r. This one uses 11 nodes compared to 12 in Fig. 2

where $\Omega \subseteq \mathbf{R}^d$ is compact. Define q as the linear functional that maps f to $\phi_f(x_0)$, where $x_0 \in \mathbf{int}\,\Omega$. By the strong maximum principle q is strictly monotone and continuous. It also satisfies $\mathbf{supp}\, q = \Omega$.

A natural choice for test functions are the n eigenfunctions of the Laplacian,

$$-\nabla^2 : H^2(\Omega) \cap \left\{ f : \frac{\partial f}{\partial \widehat{n}} = 0 \right\} \to L^2(X),$$

where $H^2(\Omega)$ is the standard Sobolev space associated with the n smallest eigenvalues. We note that $\lambda_0 = 0$ and $p^{(0)} = 1$. We take the sensitivity function r to be the eigenfunction associated with λ_n.

Now consider the specific example with the domain given by (4). Discretized versions of $p^{(i)}$ are computed as eigenvectors of the discretized Laplacian, and intermediate values between the discretized points are obtained via interpolation. Applying our method to find a quadrature of order $n = 30$, we obtain an efficient Gauss-LP quadrature with $N = 16$ nodes, as shown in Fig. 3.

5.4 Surface Integration on a Torus

We take our domain Ω to be the torus with inner and outer radii r and R, with measure

$$q(f) = \int_\Omega f(x, y, z)\, dS = \int_0^{2\pi}\int_0^{2\pi} f\,((R + r\cos\phi)\cos\theta, \\ (R + r\cos\phi)\sin\theta, r\sin\phi)\;\; d\phi d\theta.$$

In other words, $q(f)$ is the surface integral of f over the torus. We use polynomial test functions $x^p y^q z^r$ for $p + q + r < m$, so $n = m(m+1)(m+2)/6$, and the sensitivity

function $r = x^m + y^m + z^m$. The resulting Gauss-LP quadrature of order $n = 56$ with $N = 24$ nodes is shown in Fig. 4.

5.5 Solution with Infinite Support

Here we provide an example in which a solution $\mu^\star$ of (2) is supported on a continuous curve; in particular, the support is not finite. We take $\Omega = \{(x, y) \in \mathbf{R}^2 : -1 \le x, y \le 1\}$, i.e., the 2D unit square, with measure $q(f) = \int_\Omega f \, dxdy$. We use 1, x, and y as our test functions and $r = x^2$ as our sensitivity function. The LP (2) is then

$$\begin{aligned} &\text{minimize } \mu(x^2) \\ &\text{subject to } \mu(1) = 4, \\ &\qquad\qquad \mu(x) = \mu(y) = 0, \\ &\qquad\qquad \mu \ge 0, \end{aligned}$$

where $\mu \in \mathcal{M}$ is the optimization variable. Since $x^2 \ge 0$, we have $p^\star \ge 0$. So a feasible μ is a solution if $\mu(x^2) = 0$. Finally, $d\mu^\star = 2d\delta_0(x)dy$ is a solution and $\mathbf{supp}\,\mu^\star$ is the y-axis. Clearly, $|\mathbf{supp}\,\mu^\star| = \infty$.

From Sect. 3 we know that there must be another solution supported on at most 3 points. One such example is $d\mu = 4d\delta_0(x)d\delta_0(y)$, which is supported on 1 point.

6 Conclusion and Extensions

The procedure presented in this paper provides a heuristic method to obtain efficient quadratures for general integrals in multiple dimensions, a regime where the traditional theory of the Gauss quadrature breaks down. While we do not have a theoretical guarantee that the procedure will produce an efficient quadrature (except in the 1D case) we observe empirically that it often does.

We mention here a few interesting extensions of and variations on the method. The first concerns the choice of the sensitivity function r. We have observed that different sensitivity functions lead to different Gauss-LP quadratures and, indeed, Gauss-LP quadratures with different numbers of nodes. As an example, take the domain to be the 2D unit square, the test functions to be the bivariate monomials of degree less than m, and $q(f) = \int_\Omega f \, dxdy$. If we let $r = x^m + y^m$, then the solution to (2) is the product of 1D Gauss quadratures. However, if we choose $r = \cos(\frac{\pi}{2}\sqrt{(x+1)^2 + (y+1)^2})$, then we obtain a quadrature of the same order with fewer nodes, as shown in Fig. 5. Likewise, in the setting of Sect. 5.2 we find a quadrature with fewer node points with the choice $r = \exp(x^2 + y^3)$, as shown in Fig. 6. This suggests the possiblity of choosing r to minimize N, the number of node points. We do not know how to do this; moreover, it is harder to justify an unnatural choice of r than a natural one.

Another variation is to use a nonlinear objective instead of the linear objective. In particular, $\max\{\mu(f_1), \ldots, \mu(f_k)\}$ is an interesting choice because it can be given a minimax interpretation and would include $|\mu(f)| = \max\{\mu(f), \mu(-f)\}$. These problems can be solved as LPs as well.

Finally, we could consider the extension to signed measures. For signed measures it makes sense to use the objective $\|\mu\|$, which would be an infinite-dimensional analog of ℓ_1 minimization.

Acknowledgments We thank Pablo Parrilo for helpful conversations; indeed, this paper grew out of conversations with him. We also thank Paul Constantine for his helpful feedback. Ernest Ryu is supported by the Department of Energy Office of Science Graduate Fellowship Program (DOE SCGF) and by the Simons Foundation.

Appendix 1: Notation

We write C_b for the Banach space of bounded continuous functions on Ω. Write $\|\cdot\|_\infty : C_b \to \mathbf{R}$ for the supremum norm defined as

$$\|f\|_\infty \stackrel{\text{def}}{=} \sup_{x\in\Omega} |f(x)|.$$

We write $\mathcal{M}$ for the Banach space of finite signed Borel measures on Ω and for $\mu \in \mathcal{M}$ write $\mu \geq 0$ to denote that μ is unsigned. An unsigned measure μ is finite if $\mu(\Omega) < \infty$.

The *support* of $\mu \in \mathcal{M}$ is defined as

$$\mathbf{supp}\,\mu = \{x \in \Omega | \forall\, r > 0,\ \mu(B(x,r)) > 0\},$$

and $|\mathbf{supp}\,\mu|$ denotes the cardinality of $\mathbf{supp}\,\mu$ as a set.

We write $\mathcal{N}$ for the Banach space of normal signed Borel *charges* of bounded variation. A charge is a set function defined on an algebra and is like a measure except that it is only finitely (and not necessarily countably) additive. A charge is *Borel* if it is defined on the algebra generated by open sets. A charge is *normal* if $\mu(A) = \sup\{\mu(F) : F \subseteq A,\ F \text{ closed}\} = \inf\{\mu(G) : A \subseteq G,\ G \text{ open}\}$. A charge is *tight* if $\mu(A) = \sup\{\mu(K) : K \subseteq A,\ K \text{ compact}\}$. For $\mu \in \mathcal{M}$ write $\mu \geq 0$ to denote that μ is unsigned. An unsigned charge μ is of bounded variation if $\mu(\Omega) < \infty$. Integration with respect to a charge is defined similarly to Lebesgue integration [2, §14.2].

By Theorem 5, $\mathcal{N}$ is isomorphic to the dual of C_b. Thus, for any $\mu \in \mathcal{N}$ and $f \in C_b$ we can view μ as a linear functional acting on f, and we denote this action as

$$\langle f, \mu\rangle \stackrel{\text{def}}{=} \int_\Omega f\, d\mu.$$

Appendix 2: Miscellaneous Theorems

Theorem 3 *If* $\mathbf{supp}\, q = \Omega$*, a Gauss quadrature is the only quadrature that integrates* $1, x, \ldots, x^{n-1}$ *exactly with* $n/2$ *or fewer nodes [31].*

Theorem 4 *On $\Omega \subseteq \mathbf{R}^d$, every tight finite Borel charge is a measure [2, §12.1]. (Precisely speaking, the charge has a unique extension to the Borel σ-algebra that is a measure.)*

In the proof of strong duality we will encounter charges, which are generalizations of measures. Fortunately, Theorem 4 will allow us to conclude that the charge is in fact a measure.

Theorem 5 *For a domain Ω not necessarily compact, the dual space C_b^* is isomorphic to $\mathcal{N}$, the space of signed normal Borel charges of bounded variation [2, §14.3].*

Theorem 5, used in the proof of Theorem 1, is analogous to the Riesz–Markov theorem and provides an explicit representation of the dual space C_b^*.

Appendix 3: Proof of Main Results

Appendix 3.1: Proof of Theorem 1

We will call the optimization problem (1) the *primal* problem and the following problem the *dual* problem.

$$\begin{aligned} \text{maximize} \quad & \sum_{i=0}^{n-1} \nu_{i+1} \langle x^i, q \rangle \\ \text{subject to} \quad & \lambda = x^n - \sum_{i=0}^{n-1} \nu_{i+1} x^i, \\ & \lambda(x) \geq 0, \quad \text{for all } x \in \Omega, \end{aligned} \tag{5}$$

where $\nu \in \mathbf{R}^n$ is the optimization variable. We define $\mu^\star$ and $\nu^\star$ as solutions of the primal and dual problems, respectively. We write $\lambda^\star$ for the polynomial that corresponds to $\nu^\star$. Let $p^\star$ and $d^\star$ denote the optimal values of the primal and dual problems.

In the proof we first introduce a new LP, problem (6), that is similar to the original LP, problem (1), but different in that is has a larger search space. We show that problem (6) is the dual of problem (5) and that strong duality and complementary slackness holds. The nonnegative polynomial $\lambda^\star$ can have at most $n/2$ roots, and this will allow us to conclude that in fact problems (6) and (1) must share the same solution and that the solution must be a Gauss quadrature.

Proof Define

$$\psi(x) = \begin{cases} 1 & \text{if } |x| \leq 1 \\ \frac{1}{x^n} & \text{otherwise} \end{cases}$$

and the norm

$$\|f\|_{\infty,\psi} = \|f\psi\|_\infty = \sup_{x \in \Omega} |f(x)\psi(x)|.$$

Let D be the set of continuous real-valued functions f defined on Ω such that $\|f\|_{\infty,\psi} < \infty$. In other words, $D = \frac{1}{\psi} C_b$ is the space of continuous functions that grow at a rate of at most $\mathcal{O}(x^n)$.

The map $T : D \to C_b$, where $T : f \mapsto f\psi$, is an isometric lattice isomorphism between $(D, \|\cdot\|_{\infty,\psi})$ and $(C_b, \|\cdot\|_\infty)$ [2, §14.3]. Thus, $(D, \|\cdot\|_{\infty,\psi})$ is a Banach space. Moreover, since $C_b^* \cong \mathcal{N}$ by Theorem 5, the isomorphism tells us that $D^* \cong \psi\mathcal{N}$, where $\mathcal{N}$ is the Banach space of normal Borel charges of bounded variation.

Consider the following variant of the primal problem: (1)

$$\begin{aligned} &\text{minimize} && \langle x^n, \mu \rangle \\ &\text{subject to} && \langle x^i, \mu \rangle = \langle x^i, q \rangle, \quad i = 0, \ldots, n-1, \\ &&& \mu \geq 0, \end{aligned} \tag{6}$$

where $\mu \in \psi\mathcal{N} \cong D^*$ is the optimization variable. Note that μ, which used to be in $\mathcal{M}$, now resides in a larger space. Weak duality between (6) and (5) can be readily shown via standard arguments. Both primal and dual problems are feasible because $\mu = q$ and $\nu = 0$ are feasible points, and therefore $-\infty < d^\star \leq p^\star < \infty$.

Now we can apply Lagrange duality, which states: if $d^\star$ is finite (which we have shown) and if there is a strictly feasible ν, then strong duality holds, a primal solution exists, and complementary slackness holds [21, §8.6]. The point $\nu = e_1$ is strictly feasible, and this establishes strong duality and the existence of a primal solution $\mu^\star$.

We now claim that the dual problem attains the supremum, i.e., a solution $\nu^\star$ exists. We prove this in Appendix 3.2.

Next we will show that $\mu^\star$ is a measure (not just a charge) and that $\mathbf{supp}\, \mu^\star \subseteq \{x_1, x_2, \ldots, x_k\}$, where $x_1, x_2, \ldots, x_k$ are the roots of the polynomial $\lambda^\star$. Complementary slackness states that $\langle \mu^\star, \lambda^\star \rangle = 0$. Remember that $\lambda^\star \geq 0$ by definition. For any set $A \subseteq \mathbf{R} \setminus \bigcup_{i=1}^k B(x_i, \varepsilon)$, where $\varepsilon > 0$ is small, there exists a small enough $\delta > 0$ such that $\delta 1_A \leq \lambda^\star$, where 1_A is the indicator function, and we have

$$\delta\mu^\star(A) = \int_\Omega \delta 1_A \, d\mu^\star \leq \int_\Omega \lambda^\star \, d\mu^\star = 0.$$

So we conclude $\mu^\star(A) = 0$. Now by the normality of the charge $\mu^\star$,

$$\mu^\star((x_i - \varepsilon, x_i)) = \sup\{\mu^\star(F) : F \text{ closed and } F \subseteq (x_i - \varepsilon, x_i)\}.$$

However, by the previous argument, $\mu^\star(F) = 0$ for any closed F such that $F \subseteq (x_i - \varepsilon, x_i)$. Therefore, $\mu^\star((x_i - \varepsilon, x_i)) = 0$ and, by the same logic, $\mu^\star((x_i, x_i + \varepsilon)) = 0$. Thus, $\mu^\star(\mathbf{R} \setminus \{x_1, x_2, \ldots, x_k\}) = 0$, and for any measurable set A we have $\mu^\star(A) = \mu^\star(A \cap \{x_1, x_2, \ldots, x_k\})$. In particular, $\{x_1, x_2, \ldots, x_k\}$ is compact, and this establishes the tightness of $\mu^\star$. Thus, by Theorem 4, we conclude that $\mu^\star$ is a discrete measure and can have point masses only at $x_1, x_2, \ldots, x_k$.

Since $\mu^\star$ is a discrete measure,

$$\mu^\star \in \{\mu \in \mathcal{M} | \mu \text{ is feasible for the primal problem (1)}\} \subseteq \psi\mathcal{N}.$$

Therefore, the primal problem can be simplified by searching over the feasible measures in $\mathcal{M}$ and not over the entire superspace $\psi\mathcal{N}$. In other words, $\mu^\star$, the solution to problem (6), is a solution to the original problem (1).

Moreover, since $\lambda^\star$ is a nonnegative polynomial of degree n, it can have at most $n/2$ distinct roots in Ω, and therefore $|\mathbf{supp}\,\mu^\star| \leq n/2$. In other words, $\mu^\star$ is equivalent to a quadrature that integrates $1, x, x^2, \ldots, x^{n-1}$ exactly with $n/2$ or fewer nodes. Thus, by Theorem 3, we conclude that $\mu^\star$ must be the Gauss quadrature and that the solution $\mu^\star$ is unique. □

Appendix 3.2: Attainment of Dual Optimum

Lemma 1 *Let $K \subseteq \mathbf{R}^n$ be a proper cone. Assume $u_0 \in K^*$ has the property that for any $v \in K$ we have $v^T u_0 > 0$, unless $v = 0$. Then $u_0 \in \mathbf{int}\,(K^*)$.*

Proof Assume for contradiction that $u_0 \in \partial K^*$. Then there exists a nonzero separating hyperplane λ such that $\lambda^T u_0 = 0$ and $\lambda^T u \geq 0$ for any $u \in K^*$. However, this implies that $\lambda \in K^{**} = K$, and this contradicts the assumption that $v^T u_0 > 0$ for any nonzero $v \in K$. Thus we conclude that $u_0 \in \mathbf{int}\,(K^*)$. □

Theorem 6 *A solution to the dual problem (5) is attained.*

Proof Let K be the convex cone defined as

$$K = \{y \in \mathbf{R}^{n+1} | y_{n+1}x^n + y_n x^{n-1} + \cdots + y_2 x + y_1 \geq 0 \text{ for } x \in \Omega\},$$

i.e., the cone of coefficients of nonnegative polynomials.

Also, let M be the convex cone defined as

$$M = \{m \in \mathbf{R}^{n+1} : m = \left(\langle 1, \mu\rangle, \langle x, \mu\rangle, \ldots, \langle x^{n-1}, \mu\rangle, \langle x^n, \mu\rangle\right),\ \mu \geq 0\}, \quad (7)$$

i.e., the cone of possible moments. We note that $K^* = \mathbf{cl}\,M$, where $\mathbf{cl}\,M$ denotes the closure of M, and that $K^{**} = K$ as K is a proper cone [5, pp. 65–66]. The following duality argument hinges on these facts.

Let $m_0 \in \mathbf{R}^{n+1}$ be the moment vector, i.e., $(m_0)_{i+1} = q(x^i)$ for $i = 0, \ldots, n$. Consider the following problem:

$$\begin{aligned} \text{minimize} \quad & x_{n+1} \\ \text{subject to} \quad & x_i = (m_0)_i, \quad i = 1, \ldots, n, \\ & x \in K^*, \end{aligned} \quad (8)$$

where $x \in \mathbf{R}^{n+1}$ is the optimization variable. (Since $K^* = \mathbf{cl}\,M$, problem (8) is in fact equivalent to problem (1).) Lagrange duality tells us that the dual of (8) is (5) and that a dual solution $\nu^\star$ exists if (8) has a strictly feasible point (i.e., if Slater's constrain qualification holds) [5, §5.3]. We omit the Lagrange dual derivation because it is routine and involves no unexpected tricks.

Consider any $y \in K$ such that $y \neq 0$. Then

$$y^T m_0 = \int_\Omega \sum_{i=0}^{n} y_{i+1} x^i \, dq > 0$$

since by definition $y \in K$ implies $\sum_{i=0}^{n} y_{i+1} x^i \geq 0$ and since $\mathbf{supp}\, q = \Omega$. Thus, by Lemma 1, we see that $m_0 \in \mathbf{int}\,(K^*)$, i.e., m_0 is strictly feasible. Hence, we conclude that a dual solution $v^\star$ exists. □

Appendix 3.3: Proof of Theorem 2

Proof We write

$$M = \{m \in \mathbf{R}^{n+1} : m = \left(\langle p^{(0)}, \mu\rangle, \langle p^{(1)}, \mu\rangle, \ldots, \langle p^{(n-1)}, \mu\rangle, \langle r, \mu\rangle\right), \ \mu \geq 0\}, \tag{9}$$

and we will call M the *moment cone*. Also, define

$$K = \left\{(p^{(0)}(x), p^{(1)}(x), \ldots, p^{(n-1)}(x), r(x)) \in \mathbf{R}^{n+1} : x \in \Omega\right\}.$$

Since the $p^{(i)}$ and r are continuous, K is the image of a compact set under a continuous function and therefore is compact. We assume as before that $p^{(0)} = 1$ and, therefore, $0 \notin \mathbf{conv} K$, where $\mathbf{conv} K$ denotes the convex hull of K. Therefore, $\mathbf{cone}\, K$, the conical hull of K, is closed [16, §1.4].

Now we prove $\mathbf{cone}\, K = M$. By choosing μ to have point masses at a finite number of points in (9), we can produce any element in $\mathbf{cone}\, K$ and therefore $\mathbf{cone}\, K \subseteq M$. Now assume for contradiction that $\mathbf{cone}\, K \neq M$. In other words, assume there exists an $m \in M$ such that $m \notin \mathbf{cone}\, K$. Since $\mathbf{cone}\, K$ is a closed convex set, there must be a strictly separating hyperplane λ such that $\lambda^T m < 0$ and $\lambda^T n \geq 0$ for any $n \in \mathbf{cone}\, K$. However, since $m \in M$, there must exist a corresponding measure $\mu \geq 0$ that produced m in (9), i.e.,

$$m_{i+1} = \langle p^{(i)}, \mu\rangle \quad \text{for } i = 0, \ldots, n-1 \quad \text{and} \quad m_{n+1} = \langle r, \mu\rangle.$$

Therefore,

$$\lambda^T m = \left\langle \lambda_{n+1} r + \sum_{i=0}^{n-1} \lambda_{i+1} p^{(i)}, \mu \right\rangle < 0,$$

and this in particular implies that

$$\lambda_{n+1} r(x) + \sum_{i=0}^{n-1} \lambda_{i+1} p^{(i)}(x) < 0$$

for some $x \in \Omega$. However, since by construction $\lambda^T n \geq 0$ for all $n \in K \subseteq \mathbf{cone}\, K$, i.e.,

$$\lambda_{n+1} r(x) + \sum_{i=0}^{n-1} \lambda_{i+1} p^{(i)}(x) \geq 0 \quad \text{for all } x \in \Omega,$$

we have a contradiction. Therefore, $\mathbf{cone}\, K = M$ and M is closed.

Now consider the optimization problem

$$\begin{array}{ll} \text{minimize} & m_{n+1} \\ \text{subject to} & m_i = q(p^{(i)}), \quad i = 0, \ldots, n-1, \\ & m \in M, \end{array}$$

where $m \in \mathbf{R}^{n+1}$ is the optimization variable. Note that this problem is equivalent to the original problem (2). Since M is closed, so is the feasible set. Moreover, the feasible set is bounded because for any $m \in M$, the last coordinate m_{n+1} (the only one that is not fixed) is bouned since

$$|m_{n+1}| = |\mu(r)| \leq \|r\|_\infty \mu(1) = \|r\|_\infty q(1) < \infty$$

for some nonnegative measure $\mu \in \mathcal{M}$. Therefore, the feasible set is compact. Finally, the feasible set is nonempty because $m \in M$ generated by the measure q is a feasible point. Therefore, there exist an optimal $m^\star$ for the reduced problem and a $\mu^\star$ that generated $m^\star$; this $\mu^\star$ is optimal for the original problem (2).

Now, by Carathéodory's theorem on cones [29, §17], $m^\star \in \mathbf{cone}\, K = M$ can be expressed as a linear combination of at most $n+1$ vectors in K. This linear combination is equivalent to a measure with point masses at at most $n+1$ locations. In other words, $m^\star$ can be produced (in the sense of (9)) by a measure $\mu^\star$, where $|\mathbf{supp}\, \mu^\star| \leq n+1$. This $\mu^\star$ is a solution to problem (2).

Finally, we can further reduce the support of this solution. Given a solution $\mu^\star$ with finite support, we can restrict problem (2) to searching only over measures that are supported on $\mathbf{supp}\, \mu^\star$. This reduces problem (2) to a finite-dimensional LP, which always has a solution supported on n or fewer points [22, §2.4]. □

References

1. M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*. Dover Publications, Incorporated, New York, 1964.
2. C. Aliprantis and K. Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer, New York, 3rd edition, 2006.
3. E. Anderson and P. Nash. *Linear Programming in Infinite-dimensional Spaces: Theory and Applications*. Wiley, New York, 1987.
4. E. Anderson and A. Philpott. *Infinite Programming: Proceedings of an International Symposium on Infinite Dimensional Linear Programming, Churchill College, Cambridge, United Kingdom, September 7–10, 1984*. Springer-Verlag, New York, 1985.
5. S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, 2004.

6. E. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.
7. S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1999.
8. P. Davis. *A construction of nonnegative approximate quadratures. Mathematics of Computation*, 21(100):578–582, 1967.
9. P. Davis and P. Rabinowitz. *Methods of Numerical Integration*. Academic Press, Orlando, 2nd edition, 1984.
10. A Fiacco and K. Kortanek. *Semi-infinite Programming and Applications: An International Symposium, Austin, Texas, September 8–10, 1981*. Springer-Verlag, New York, 1983.
11. C. Gauss. Methodus nova integralium valores per approximationem inveniendi. *Commentationes Societatis Regiae Scientarium Gottingensis Recentiores*, 3:39–76, 1814.
12. A. Glaser, X. Liu, and V. Rokhlin. A fast algorithm for the calculation of the roots of special functions. *SIAM Journal on Scientific Computing*, 29(4):1420–1438, 2007.
13. M. Goberna and M. López. *Linear Semi-infinite Optimization*. John Wiley, 1998.
14. G. Golub and J. Welsch. Calculation of gauss quadrature rules. *Mathematics of Computation*, 23(106):221–230, 1969.
15. P. Hammer and A. Stroud. Numerical evaluation of multiple integrals II. *Mathematics of Computation*, 12:272–280, 1958.
16. J. Hiriart-Urruty and C. Lemarechal. *Convex Analysis and Minimization Algorithms I: Fundamentals*. Springer, New York, October 1993.
17. S. Karlin and W. Studden. *Tchebycheff Systems: With Applications in Analysis and Statistics*. Interscience Publishers, New York, 1966.
18. M. Krein. The ideas of P. L. Chebyshev and A. A. Markov in the theory of limiting values of integrals and their further development. *American Mathematical Society Translations (Series 2)*, 12:1–122, 1959.
19. V. Krylov. *Approximate Calculation of Integrals*. Macmillan, New York, 1962.
20. J. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11:796–817, 2001.
21. D. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, New York, 1967.
22. D. Luenberger and Y. Ye. *Linear and Nonlinear Programming*. Springer, New York, third edition, 2008.
23. J. Ma, V. Rokhlin, and S. Wandzura. Generalized Gaussian quadrature rules for systems of arbitrary functions. *SIAM Journal on Numerical Analysis*, 33(3):971–996, 1996.
24. A. Markov. On the limiting values of integrals in connection with interpolation. *Zapiski Imperatorskoj Akademii Nauk po Fiziko-matematiceskomu Otdeleniju*, 5:146–230, 1898.
25. G. Murty and S. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39(2):117–129, 1987.
26. J. Nocedal and S. Wright. *Numerical Optimization*. Springer, New York, 2nd edition, 2006.
27. W. Peirce. Numerical integration over the planar annulus. *Journal of the Society for Industrial and Applied Mathematics*, 5(2):66–73, 1957.
28. J. Powell. *Approximation Theory and Methods*. Cambridge University Press, Cambridge, 1981.
29. R. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1996.
30. A. Stroud. Quadrature methods for functions of more than one variable. *Annals of the New York Academy of Sciences*, 86(3):776–791, 1960.
31. A. Stroud and D. Secrest. *Gaussian Quadrature Formulas*. Prentice-Hall, Englewood Cliffs, 1966.
32. E. Süli and D. Mayers. *An Introduction to Numerical Analysis*. Cambridge University Press, Cambridge, 2003.
33. R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society (Series B)*, 58:267–288, 1996.
34. B. Vioreanu. *Spectra of Multiplication Operators as a Numerical Tool*. PhD thesis, Yale University, 2012.
35. H. Xiao and Z. Gimbutas. A numerical algorithm for the construction of efficient quadrature rules in two and higher gimensions. *Computational Mathematics with Applications*, 59(2), 2010.