Graham Alldredge
January 21, 2015

Let $\mathcal{R}_{\mathbf{b}}$ be the realizable set

$$\mathcal{R}_{\mathbf{b}} := \left\{ \mathbf{u} = (u_0, \dots, u_n)^T \in \mathbb{R}^{n+1} : \exists f \geq 0, \text{ such that } \langle \mathbf{b} f \rangle = \mathbf{u} \right\}, \quad (1)$$

where $\mathbf{b} = \mathbf{b}(\mu) \in \mathbb{R}^{n+1}$ and $\langle \cdot \rangle$ indicates integration of each component of its argument over $\mu \in [-1, 1]$, and let $\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}$ be its approximation by a quadrature $\mathcal{Q}$:

$$\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}} := \left\{ \mathbf{u} \in \mathbb{R}^{n+1} : \exists f_i \geq 0, \text{ such that } \sum_{i=1}^{Q} w_i \mathbf{b}(\mu_i) f_i = \mathbf{u} \right\}, \quad (2)$$

where $\{w_i\}$ are the (positive) quadrature weights and $\{\mu_i\}$ are the quadrature nodes. For convenience, we are defining $\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}$ to be closed, and so for consistency, the f considered in the definition of $\mathcal{R}$ should be either in $L^1(d\mu)$ or an atomic distribution. This ensures that $\mathcal{R}_{\mathbf{b}}$ is also closed and that $\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}} \subset \mathcal{R}_{\mathbf{b}}$.

For simplicity, let us assume that $b_0(\mu) \equiv 1$. Clearly $\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}\big|_{u_0=1}$ (i.e. where $1 = u_0 = \sum w_i f_i$) is the convex polytope $\text{co}\{\mathbf{b}_1(\mu_i)\}_{i=1}^{Q}$, where

$$\mathbf{b}_1(\mu) = (b_1(\mu), \dots, b_n(\mu))^T \quad (3)$$

simply removes the constant component b_0. Clearly this is a compact subset of $\mathbb{R}^n$. As a convex polytope, $\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}\big|_{u_0=1}$ has a half-space representation, that is there exist A and b such that $\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}\big|_{u_0=1} = \{\mathbf{u}_1 : A\mathbf{u}_1 \leq b\}$, where $\mathbf{u}_1 = (u_1, u_2, \dots, u_n)^T$. The matrix A has a row for every facet of the convex hull, and the vector b has just as many elements. The matrix A has a useful property:

Lemma 1. *If $A\mathbf{u}_1 \leq 0$, then $\mathbf{u}_1 = 0$.*

Proof. If there exists a $\mathbf{w}_1 \neq 0$ such that $A\mathbf{w}_1 \leq 0$, then for any $\mathbf{v}_1$ such that $A\mathbf{v}_1 \leq b$, we have $A(\mathbf{v}_1 + \alpha \mathbf{w}_1) \leq b$ for every $\alpha \geq 0$. But $\{\mathbf{u}_1 : A\mathbf{u}_1 \leq b\}$ is equal to the convex polytope $\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}\big|_{u_0=1}$ which is a compact set, so we have a contradiction. □

Since the description of $\mathcal{R}$ is complex and not very well understood when the integration domain is more than one-dimensional, the set $\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}$ and its half-space representation may indeed be a practical alternative. The main

drawback to using $\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}$, however, is it often has a very large number of facets. In this note we use a result from the study of convex polytopes to give exactly how many facets are in $\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}$ when $\mathbf{b}$ contains the monomials or the mixed-moment basis functions.

To apply the realizability limiter, we need to use the half-space representation of $\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}\big|_{u_0\leq 1}$. We first show that the number of facets of $\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}\big|_{u_0\leq 1}$ is only one more than the number of facets of $\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}\big|_{u_0=1}$, thereby allowing us to focus on $\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}\big|_{u_0=1}$ in the sequel.

Lemma 2. *If A and b define the half-space representation of $\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}\big|_{u_0=1}$, then a half-space representation of $\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}\big|_{u_0\leq 1}$ is:*

$$\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}\big|_{u_0\leq 1} = \left\{\mathbf{u} = \begin{pmatrix} u_0 \\ \mathbf{u}_1 \end{pmatrix} : \begin{pmatrix} 1 & 0 \\ -b & A \end{pmatrix} \begin{pmatrix} u_0 \\ \mathbf{u}_1 \end{pmatrix} \leq \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right\}. \tag{4}$$

Proof. To see this, let the proposed half-space representation on the right-hand side be $\mathcal{H}$. Using the fact that $\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}\big|_{u_0=\rho} = \{\mathbf{u}_1 : A\mathbf{u} \leq \rho b\}$, it is straightforward to show that $\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}\big|_{u_0\leq 1} \subseteq \mathcal{H}$. To show $\mathcal{H} \subseteq \mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}\big|_{u_0\leq 1}$ requires a bit more care.

We first need to show that there does not exist any $\mathbf{u}$ with $u_0 < 0$ in $\mathcal{H}$. Clearly $u_0 \leq 1$, so $\mathbf{u} \in \mathcal{H}$ only if it satisfies the last rows of the inequalities in the right-hand side of (4), namely if $A\mathbf{u}_1 \leq u_0 b$. It is easier to work with coordinates centered around an interior point, so let $\mathbf{v}_1$ be such that $A\mathbf{v}_1 < b$. Now for some $u_0 < 0$ we consider two cases: either when the moments $\mathbf{u}_1/|u_0|$ are in $\mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}\big|_{u_0=1}$ or when they are not:

- In the first case, since $|u_0| = -u_0$, in the translated coordinates we have $A(\mathbf{u}_1 - \mathbf{v}_1) \leq -u_0(b - A\mathbf{v}_1)$. If the desired inequality also holds, namely $A(\mathbf{u}_1 - \mathbf{v}_1) \leq u_0(b - A\mathbf{v}_1)$, then by adding these inequalities we have $2A(\mathbf{u}_1 - \mathbf{v}_1) \leq 0$. By Lemma 1, $\mathbf{u}_1 = \mathbf{v}_1$. But inserting this into the desired inequality gives $0 \leq u_0(b - A\mathbf{v}_1)$. Dividing by u_0 (which is nonpositive) and rearranging gives $A\mathbf{v}_1 \geq b$ which contradicts our original assumption on $\mathbf{v}_1$.

- Otherwise, there exists some i such that $\mathbf{a}_i^T(\mathbf{u}_1 - \mathbf{v}_1) > -u_0(b_i - A\mathbf{v}_i)$ (where $\mathbf{a}_i^T$ is the i-th row of A and b_i is the corresponding element of b). But since $A\mathbf{v}_i < b_i$ and $u_0 \leq 0$, we have $-u_0(b_i - A\mathbf{v}_i) \geq u_0(b_i - A\mathbf{v}_i)$, so altogether we have $\mathbf{a}_i^T(\mathbf{u}_1 - \mathbf{v}_1) > u_0(b_i - A\mathbf{v}_i)$. Therefore here $\mathbf{u}$ cannot be in $\mathcal{H}$.

Therefore, for any $\mathbf{u} \in \mathcal{H}$, we have either $\mathbf{u} = 0$ or $u_0 > 0$. For the former, clearly $\mathbf{u} = 0 \in \mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}\big|_{u_0 \leq 1}$. In the latter case, when $u_0 > 0$, $\mathbf{u} \in \mathcal{H}$ implies $u_0 \in (0, 1]$ and $\mathbf{u}_1/u_0 \in \mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}\big|_{u_0=1}$, which together gives $\mathbf{u} \in \mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}\big|_{u_0 \leq 1}$. Thus we have $\mathcal{H} \subseteq \mathcal{R}_{\mathbf{b}}^{\mathcal{Q}}\big|_{u_0 \leq 1}$. □

For the monomial case we let $n = N$ and $\mathbf{b}(\mu) = \mathbf{p}(\mu) = (1, \mu, \ldots, \mu^N)^T$. The convex polytope $\mathcal{R}_{\mathbf{p}}^{\mathcal{Q}}\big|_{u_0=1} \subset \mathbb{R}^N$ which contains only the normalized moments is known as the *cyclic polytope* and plays a special role in the study of convex polytopes. The Upper Bound Theorem states that for a given number of vertices in a given dimension, the cyclic polytope has the maximum number of facets. The Dehn-Sommerville equations are then used to show that the number of facets is

$$C(N, Q) = \binom{Q - \lfloor \frac{1}{2}(N+1) \rfloor}{Q - N} + \binom{Q - \lfloor \frac{1}{2}(N+2) \rfloor}{Q - N} \tag{5}$$

for $Q > N > 1$. We note that this holds independent of the choice of quadrature nodes $\{\mu_i\}$. Since $\mathcal{R}_{\mathbf{p}}^{\mathcal{Q}}\big|_{u_0=1}$ has $C(N, Q)$ facets, there exists a half-space representation such that $A \in \mathbb{R}^{C(N,Q) \times N}$ and $b \in \mathbb{R}^{C(N,Q)}$. Unpacking the definition of the binomial coefficient we can see that for fixed, even N, we have $C(N, Q) = \mathcal{O}(Q^{N/2})$, and for fixed, odd N we have $C(N, Q) = \mathcal{O}(Q^{(N-1)/2})$.

The mixed-moment case, we let $n = 2N$ and

$$\mathbf{b}(\mu) = \mathbf{m}(\mu) = \left(1, \mu_+, \ldots, \mu_+^N, \mu_-, \ldots, \mu_-^N\right)^T, \tag{6}$$

where $\mu_+ = \max(\mu, 0)$ and $\mu_- = \min(\mu, 0)$. We can write

$$\{\mathbf{m}_1\}_{i=1}^{Q} = \left\{ \left\{ \begin{matrix} \mathbf{p}_1(\mu_i) \\ 0 \end{matrix} \right\}_{\mu_i \geq 0}, \left\{ \begin{matrix} 0 \\ \mathbf{p}_1(\mu_i) \end{matrix} \right\}_{\mu_i \leq 0} \right\}. \tag{7}$$

A half-space representation for $\mathcal{R}_{\mathbf{m}}^{\mathcal{Q}}\big|_{u_0=1}$ can be derived using the half-space representations from the full-moment case.

Claim 1. *Let $A_\pm$ and $b_\pm$ define half-space representations for the convex polytopes formed by the basis functions on the positive and negative subintervals respectively:*

$$\operatorname{co}\{\mathbf{p}_1(\mu_i)\}_{\mu_i \geq 0} = \{\mathbf{u}_{1+} : A_+ \mathbf{u}_{1+} \leq b_+\}, \tag{8a}$$

$$\operatorname{co}\{\mathbf{p}_1(\mu_i)\}_{\mu_i \leq 0} = \{\mathbf{u}_{1-} : A_- \mathbf{u}_{1-} \leq b_-\}. \tag{8b}$$

We assume $b_\pm \geq 0$ component-wise.[1] *Then a half-space representation for $\mathcal{R}_{\mathbf{m}}^{\mathcal{Q}}\big|_{u_0=1}$ is given by*

$$\mathcal{R}_{\mathbf{m}}^{\mathcal{Q}}\big|_{u_0=1} = \mathcal{H} := \{\mathbf{u}_1 : A\mathbf{u}_1 \leq b\}, \tag{9}$$

where

$$A = \begin{pmatrix} A_+ & 0 \\ 0 & A_- \\ \vdots & \vdots \\ b_{+i}^{-1}\mathbf{a}_{+i}^T & b_{-j}^{-1}\mathbf{a}_{-j}^T \\ \vdots & \vdots \end{pmatrix} \quad \textit{and} \quad b = \begin{pmatrix} b_+ \\ b_- \\ \vdots \\ 1 \\ \vdots \end{pmatrix}; \tag{10}$$

where the last rows of the half-space conditions express that $b_{+i}^{-1}\mathbf{a}_{+i}^T\mathbf{u}_{1+} + b_{-j}^{-1}\mathbf{a}_{-j}^T\mathbf{u}_{1-} \leq 1$ for any pair $\{i,j\}$ such that neither b_{+i} nor b_{-j} is equal to zero.

Proof. We first show $\mathcal{R}_{\mathbf{m}}^{\mathcal{Q}}\big|_{u_0=1} \subseteq \mathcal{H}$. If $\mathbf{u}_1 \in \mathcal{R}_{\mathbf{m}}^{\mathcal{Q}}\big|_{u_0=1}$, then there exists $\{\lambda_{+i}\}$ and $\{\lambda_{-i}\}$ such that

$$\mathbf{u}_{1+} = \sum_{\mu_i \geq 0} \lambda_{+i}\mathbf{p}_1(\mu_i), \tag{11a}$$

$$\mathbf{u}_{1-} = \sum_{\mu_i \leq 0} \lambda_{-i}\mathbf{p}_1(\mu_i), \text{ and} \tag{11b}$$

$$1 = \sum_{\mu_i \geq 0} \lambda_{+i} + \sum_{\mu_i \leq 0} \lambda_{-i}. \tag{11c}$$

Now let $\lambda_\pm := \sum \lambda_{\pm i}$ respectively. Therefore $\mathbf{u}_{1+}/\lambda_+$ and $\mathbf{u}_{1-}/\lambda_-$ belong to the sets in (8) respectively, and therefore

$$A_\pm \mathbf{u}_{1\pm} \leq \lambda_\pm b_\pm \leq b_\pm \tag{12}$$

respectively, since $b_\pm \geq 0$, thus satisfying the first two block rows in $\mathcal{H}$. For every row such that $b_{\pm i} \neq 0$, we also have $\mathbf{a}_{\pm i}^T\mathbf{u}_{1\pm} \leq \lambda_\pm b_{\pm i}$. Also from the rows of (12) we have

$$b_{\pm i}^{-1}\mathbf{a}_{\pm i}^T\mathbf{u}_{1\pm} \leq \lambda_\pm \tag{13}$$

[1]Here we are using the fact that the subintervals for the mixed-moments are joined exactly at $\mu = 0$ and assume furthermore that this point is a quadrature node. This is indeed a reasonable assumption, since even in MM_1, a delta function can form at $\mu = 0$.

for any such row, respectively. Now summing these two inequalities for any combination of these rows gives the last block row in $\mathcal{H}$, so we have $\mathcal{R}^{\mathcal{Q}}_{\mathbf{m}}\big|_{u_0=1} \subseteq \mathcal{H}$.

Now assume $\mathbf{u}_1 \in \mathcal{H}$. From the first two block rows of A and b, we have $A_\pm \mathbf{u}_{1\pm} \leq b_\pm$, so $\lambda_\pm := \max_i \mathbf{a}_{\pm i}^T \mathbf{u}_{1\pm}/b_{\pm i} \leq 1$, where the maxima are only taken over rows such that $b_{\pm i} > 0$, respectively. Now if $\lambda_+ \leq 0$, then by Lemma 1 we have $\mathbf{u}_{1+} = 0$. Then using $A_-\mathbf{u}_- \leq b_-$, we have that $(0, \mathbf{u}_{1-}) \in \mathcal{R}^{\mathcal{Q}}_{\mathbf{m}}\big|_{u_0=1}$. A similar argument holds when $\lambda_- \leq 0$.

Otherwise, $\lambda_\pm \in (0, 1]$, and from their definitions we have $A_\pm \mathbf{u}_{1\pm} \leq \lambda_\pm b_\pm$. From one of the last rows of $A\mathbf{u}_1 \leq b$ we also have $\lambda_+ + \lambda_- \leq 1$, so $A_-\mathbf{u}_{1-} \leq \lambda_- b_- \leq (1 - \lambda_+)b_-$. Thus

$$\frac{\mathbf{u}_{1+}}{\lambda_+} \in \mathrm{co}\{\mathbf{p}_1(\mu_i)\}_{\mu_i \geq 0} \quad \text{and} \quad \frac{\mathbf{u}_{1-}}{1-\lambda_+} \in \mathrm{co}\{\mathbf{p}_1(\mu_i)\}_{\mu_i \leq 0}, \tag{14}$$

which also shows that $\mathbf{u}_1 \in \mathcal{R}^{\mathcal{Q}}_{\mathbf{m}}\big|_{u_0=1}$, and we conclude $\mathcal{H} \subseteq \mathcal{R}^{\mathcal{Q}}_{\mathbf{m}}\big|_{u_0=1}$. □

Remark. The assumption $b_\pm \geq 0$ is indeed *with* some loss of generality: a simple translation of the vertices is trickier than it sounds here because the translation must be applied to the rows of zeros in the vertices.

The number of rows such that $b_{\pm i} = 0$ is equal to the number of facets including the vertex corresponding to the quadrature point at $\mu = 0$. These facets can be more generally described as those containing the vertex corresponding to the first quadrature point, when the quadrature points are arranged in increasing order. The number of such facets can be computed using Gale's evenness condition:

$$C_0(N, Q) = \begin{cases} \binom{Q - (N+1)/2}{Q - N} + \binom{Q - (N+3)/2}{Q - N} & \text{if } N \text{ is odd,} \\ 2\binom{Q - (N+2)/2}{Q - N} & \text{if } N \text{ is even.} \end{cases} \tag{15}$$

(See Theorem 13.6 and Exercise 13.1 in the Brønsted book). Thus, when N is odd, $C_0(N, Q) = \mathcal{O}(Q^{(N-1)/2})$, and when N is even, $C_0(N, Q) = \mathcal{O}(Q^{N/2-1})$. In both cases, the order of $C(N, Q) - C_0(N, Q)$ is the same as $C(N, Q)$.

If we let $C_\pm$ denotes the number of rows of $A_\pm$ respectively, this representation gives $C_+ + C_- + C_+C_-$ as an upper-bound on the number of facets in $\mathcal{R}^{\mathcal{Q}}_{\mathbf{m}}\big|_{u_0=1}$. For now I am unable to show that none of the inequalities in this

half-space representation are redundant, so the true number of facets may be smaller.

To compare the full-moment and mixed-moment cases for the same number of degrees of freedom, we consider the full-moment case of order N, for N even, and the mixed-moment case of order $N/2$. Let us assume that we use a quadrature set which includes $\mu = 0$ and has $Q/2$ points over both $\mu \geq 0$ and $\mu \leq 0$, for a total of $Q-1$ points (since the point at $\mu = 0$ is counted twice). Then the number of facets in the full-moment case is $C(N, Q-1)$ while in the mixed-moment case it is on the order of $C(N/2, Q/2)^2$. When $N/2$ is odd we have $\mathcal{O}(Q^{N/2-1})$, which is one order less than in the full-moment case.

When $N/2$ is even, our half-space representation for the mixed-moment case has $\mathcal{O}(Q^{N/2})$ facets. This is the same order as the full-moment case, but we can show that the leading-order coefficient is smaller in the mixed-moment case, thereby showing that the number of facets in the mixed-moment case is at least asymptotically smaller. Let $N = 4n$. For the full-moment case we have

$$C(4n, Q-1) = \frac{(Q-1-2n)!}{(Q-1-4n)!(2n)!} + \mathcal{O}(Q^{2n-1}) = \frac{1}{(2n)!}Q^{2n} + \mathcal{O}(Q^{2n-1}). \quad (16)$$

In the mixed-moment case, we can ignore the facets lost where $b_{\pm i} = 0$ (because the number of such terms is of a lower order), so the coefficient of the leading-order term is given by computing

$$C(2n, Q/2)^2 = \left(\frac{\left(\frac{Q}{2} - n\right)!}{\left(\frac{Q}{2} - 2n\right)!n!} + \mathcal{O}(Q^{n-1}) \right)^2 = \frac{1}{(2^n n!)^2}Q^{2n} + \mathcal{O}(Q^{2n-1}). \quad (17)$$

Then the ratio of the $2n$-th order term of $C(2n, Q/2)^2$ over the $2n$-th order term of $C(4n, Q)$ is $(2n)!/(2^n n!)^2$, which is bounded by $1/2$ and in fact goes to zero as $n \to \infty$.

I have made the following observations when comparing this analysis with the results of `convhulln` in Matlab:

- For $N \in \{3, 4, \ldots 11\}$ and $Q = \{15, 20, \ldots 45\}$, the number of facets computed by `convhulln` for $\mathcal{R}_{\mathbf{p}}^{\mathcal{Q}}\big|_{u_0=1}$ is always exactly right!
- However, in the full-moment case, the number of facets `convhulln` computes for $\mathcal{R}_{\mathbf{p}}^{\mathcal{Q}}\big|_{u_0 \leq 1}$ is much higher than predicted here. Oddly enough, the number of facets is often equal to the $C(N+1, Q+1)$.

- The half-space representation given here for the mixed-moment case has often has fewer facets than what `convhulln` returns.