Supplement to "Sharp Pair Selection for Mean Regression"

Pahan Dewasurendra
Johns Hopkins University

A A Standalone Atomic Proof

We restate the geometric result in a form that permits repeated support points and zero masses.

Theorem 1. *Let $p \in \mathbb{R}_+^m$ satisfy $\sum_i p_i = 1$, and let $x_1, \ldots, x_m$ lie in a real Hilbert space. If $\sum_i p_i x_i = 0$, then some i, j satisfy*

$$\left\| \frac{x_i + x_j}{2} \right\|^2 \le \max\left\{ \frac{1}{3}, \frac{m-2}{2(m-1)} \right\} \sum_i p_i \|x_i\|^2.$$

Zero-mass points may be removed. The constant is sharp when at most m positive masses are allowed.

Proof. Remove zero masses and let m' be the remaining support size. The claimed constant is nondecreasing in support size, so it is enough to prove the result with $m = m'$. Scale the second moment to one.

If $p_j \ge 1/4$, assign probability $(4p_j - 1)/3$ to (j, j) and probability $4p_i/3$ to (i, j) for each $i \ne j$. These probabilities sum to one. Direct expansion gives

$$\begin{aligned} 3\mathbb{E} \left\| \frac{x_I + x_J}{2} \right\|^2 &= (4p_j - 1)\|x_j\|^2 + \sum_{i \ne j} p_i \|x_i + x_j\|^2 \\ &= \sum_i p_i \|x_i\|^2 + 4p_j \|x_j\|^2 - \|x_j\|^2 \\ &\quad + (1 - 4p_j)\|x_j\|^2 = 1. \end{aligned}$$

The middle line uses $\sum_{i \ne j} p_i x_i = -p_j x_j$.

Suppose next that every $p_i < 1/4$. Put

$$S = \sum_i \frac{p_i}{1 - 4p_i}, \qquad C = \frac{S+1}{3S+1},$$
$$a_i = \frac{C p_i}{(1+S)(1-4p_i)}.$$

The identity

$$\sum_i \frac{p_i^2}{1 - 4p_i} = \frac{S-1}{4}$$

implies

$$\sum_i a_i = \frac{CS}{1+S}, \qquad \sum_i p_i a_i = \frac{C(S-1)}{4(1+S)}.$$

For $i < j$, let $w_{ij} = 4(p_i a_j + p_j a_i)$. Then

$$\sum_{i<j} w_{ij} = 4\left(\sum_i a_i - \sum_i p_i a_i \right) = \frac{C(3S+1)}{S+1} = 1.$$

Let $G_{ij} = \langle x_i, x_j \rangle$. If $q = (e_i + e_j)/2$ with probability w_{ij} and $Q = \mathbb{E}[qq^\top]$, then

$$Q_{ij} = p_i a_j + a_i p_j \quad (i \ne j).$$

Also,

$$\begin{aligned} Q_{ii} &= p_i \sum_{j \ne i} a_j + a_i \sum_{j \ne i} p_j \\ &= p_i A + a_i(1 - 2p_i) = C p_i + 2 p_i a_i, \end{aligned}$$

where $A = \sum_i a_i$ and the last equality follows from $a_i(1 - 4p_i) = p_i(C - A)$. Thus

$$Q = C \operatorname{diag}(p) + p a^\top + a p^\top.$$

Since $Gp = 0$,

$$\mathbb{E} \left\| \frac{x_I + x_J}{2} \right\|^2 = \operatorname{tr}(GQ) = C \sum_i p_i \|x_i\|^2 = C.$$

Finally, convexity of $t/(1 - 4t)$ gives

$$S \ge m \frac{1/m}{1 - 4/m} = \frac{m}{m-4}.$$

Because $(S+1)/(3S+1)$ decreases in S, this implies $C \le (m-2)/(2(m-1))$. At least one pair is no worse than its expected cost.

For sharpness, the two-point distribution with masses $3/4$ and $1/4$ gives $1/3$. A uniform regular $(m-1)$-simplex gives $(m-2)/(2(m-1))$. Taking the larger proves the result. $\square$

B Matrix Certificate

The pair distributions can be checked without reference to a particular geometry. Let p be a positive probability vector and $\Sigma_p = \operatorname{diag}(p) - pp^\top$. An empirical pair frequency is a vector q of the form e_i or $(e_i + e_j)/2$.

Proposition 2. *In the dominant case, the star distribution satisfies*

$$\mathbb{E}[(q-p)(q-p)^\top] = \frac{1}{3}\Sigma_p.$$

In the diffuse case, the complete-graph distribution satisfies, on the quotient by the centered direction, the exact constant $C = (S+1)/(3S+1)$.

Proof. For the star distribution centered at j, direct calculation gives

$$\mathbb{E}q = \frac{2}{3}p + \frac{1}{3}e_j.$$

Expanding the second moment then yields the first identity. For the diffuse distribution, the main paper proves

$$\mathbb{E}[qq^\top] = C\,\mathrm{diag}(p) + pa^\top + ap^\top.$$

If G is any positive semidefinite Gram matrix with $Gp = 0$, the last two terms vanish under the trace pairing. Also $\mathrm{tr}(Gpp^\top) = 0$. Hence

$$\mathrm{tr}\left(G\mathbb{E}[(q-p)(q-p)^\top]\right) = C\,\mathrm{tr}(G\Sigma_p).$$

This is the quotient identity. □

The star formula is a Loewner-order certificate. The diffuse formula is tailored to centered geometries, which is exactly the class required after the mean is translated to zero.

C Reduction to Intrinsic Dimension

For completeness, let $D = \{z_1, \ldots, z_N\} \subset \mathbb{R}^d$ have mean μ. Translate to $u_i = z_i - \mu$ and let their span have dimension r. Consider the linear program

$$\min_{p \geq 0} \quad \sum_i p_i \|u_i\|^2$$
$$\text{subject to} \quad \sum_i p_i = 1, \qquad \sum_i p_i u_i = 0.$$

The uniform vector is feasible. A basic optimal solution has at most $r+1$ positive entries. Its objective is no larger than the full empirical variance. Applying Theorem 1 to those entries produces a pair from D with centered squared midpoint at most

$$\max\left\{\frac{1}{3}, \frac{r-1}{2r}\right\} \frac{1}{N}\sum_i \|z_i - \mu\|^2.$$

The bias-variance identity converts this into the risk ratio in the main paper.

D Equality for the Diffuse Extremizer

Assume $m \geq 5$ and a distribution on exactly m atoms attains $C_m = (m-2)/(2(m-1))$. It cannot have a mass at least $1/4$, since the star certificate would give the strictly smaller constant $1/3$. Jensen's inequality in the diffuse proof must be tight, so $p_i = 1/m$ for every i. The complete-graph pair distribution is then uniform over all distinct pairs.

Normalize the variance to one. Every distinct pair must attain the common minimum C_m, since their average equals C_m. Let $b_i = \|x_i\|^2$. For $i \neq j$,

$$\langle x_i, x_j \rangle = 2C_m - \frac{b_i + b_j}{2}.$$

Using $\sum_j x_j = 0$ and $\sum_j b_j = m$ gives

$$0 = \left\langle x_i, \sum_j x_j \right\rangle = \frac{m-4}{2}(1 - b_i).$$

Thus $b_i = 1$ and every off-diagonal inner product is $-1/(m-1)$. The Gram matrix is that of a regular simplex. This proves the equality statement in the main paper.

E Qualitative Stability

We justify the stability statement in the main paper. Fix $m \geq 5$ and consider a sequence of diffuse mass vectors $p^{(k)}$ and centered, unit-variance Gram matrices $G^{(k)}$. Suppose their best pair costs converge to $C_m = (m-2)/(2(m-1))$.

The exact diffuse certificate gives

$$\min_{i,j} \frac{G_{ii}^{(k)} + G_{jj}^{(k)} + 2G_{ij}^{(k)}}{4} \leq C(p^{(k)}) \leq C_m.$$

The left side converges to C_m, so $C(p^{(k)}) \to C_m$. Strict convexity in the Jensen step forces $p^{(k)} \to (1/m, \ldots, 1/m)$, after a fixed labeling. The masses are therefore bounded away from zero. Unit weighted trace then bounds every diagonal entry of $G^{(k)}$. Positive semidefiniteness and Cauchy–Schwarz bound every off-diagonal entry, so the Gram matrices are precompact.

For the diffuse certificate, every distinct-pair weight converges to $2/(m(m-1))$. Its weighted average pair cost equals $C(p^{(k)})$, while every pair cost is bounded below by the best pair cost. Hence every distinct-pair cost converges to C_m. Any subsequential Gram limit is centered under uniform masses, has unit weighted trace, and has all distinct midpoint costs equal to C_m. The equality argument in Section D identifies the limit as the regular-simplex Gram matrix. Thus the full sequence converges to that matrix, up to relabeling and isometry.