Supplemental Materials for "Finite-Sample Optimal Estimation and Inference on Average Treatment Effects Under Unconfoundedness"

Timothy B. Armstrong[*]
Yale University

Michal Kolesár[†]
Princeton University

December 17, 2018

D Proofs of auxiliary Lemmas and additional details

D.1 Proof of Lemma A.2

We will show that the constraint $f^*(x_i,1) \leq f^*(x_j,1) + \|x_i - x_j\|_{\mathcal{X}}$ holds for all $i,j \in \{1,\dots,n\}$. The argument that $f^*(x_i,0) \leq f^*(x_j,0) + \|x_i - x_j\|_{\mathcal{X}}$ holds for all $i,j \in \{1,\dots,n\}$ is similar and omitted. We assume, without loss of generality, that the observations are ordered so that $d_j = 0$ for $j = 1,\dots,n_0$ and $d_i = 1$ for $i = n_0+1,\dots,n$. Observe that the bias can be written as

$$\sum_{i=n_0+1}^{n}(k(x_i,1)-w(1))f(x_i,1) - \sum_{j=1}^{n_0} w(0)f(x_j,1) \\ + \sum_{j=1}^{n_0}(k(x_j,0)+w(0))f(x_j,0) + \sum_{i=n_0+1}^{n} w(1)f(x_i,0).$$

If $k(x_i,1) = w(1)$ for $i \in \{n_0+1,\dots,n\}$, we can set $f^*(x_i,1) = \min_{j\in\{1,\dots,n_0\}}\{f^*(x_j,1)+\|x_i-x_j\|_{\mathcal{X}}\}$ without affecting the bias, so that we can without loss of generality assume that (24) holds for all $i \in \{n_0+1,\dots,n\}$ and all $j \in \{1,\dots,n_0\}$.

If $w(0) = 0$, then the assumptions on k imply $k(x_i,1) = w(1)$ for $i > n_0$, and the value of $f(\cdot,1)$ doesn't affect the bias. If $w(0) > 0$, then for each $j \in \{1,\dots,n_0\}$, at least one of the constraints $f^*(x_i,1) \leq f^*(x_j,1) + \|x_i - x_j\|_{\mathcal{X}}$, $i \in \{n_0+1,\dots,n\}$, must bind, otherwise we could decrease $f^*(x_j,1)$ and increase the value of the objective function. Let $i(j)$ denote the index of one of the binding constraints (picked arbitrarily), so that $f^*(x_{i(j)},1) = f^*(x_j,1) + \|x_{i(j)} - x_j\|_{\mathcal{X}}$. We need to

[*]email: timothy.armstrong@yale.edu
[†]email: mkolesar@princeton.edu

show that the constraints

$$
\begin{aligned}
&f^*(x_i,1) \leq f^*(x_{i'},1) + \|x_i - x_{i'}\|_{\mathcal{X}} && i,i' \in \{n_0+1,\dots,n\}, && \text{(S1)}\\
&f^*(x_j,1) \leq f^*(x_{j'},1) + \|x_j - x_{j'}\|_{\mathcal{X}} && j,j' \in \{1,\dots,n_0\}, && \text{(S2)}\\
&f^*(x_j,1) \leq f^*(x_i,1) + \|x_i - x_j\|_{\mathcal{X}} && j \in \{1,\dots,n_0\},\ i \in \{n_0+1,\dots,n\}. && \text{(S3)}
\end{aligned}
$$

are all satisfied. If (S1) doesn't hold for some (i,i'), then by triangle inequality, for all $j \in \{1,\dots,n_0\}$,

$$f^*(x_{i'},1) + \|x_i - x_{i'}\|_{\mathcal{X}} < f(x_i,1) \leq f^*(x_j,1) + \|x_i - x_j\|_{\mathcal{X}} \leq f^*(x_j,1) + \|x_i - x_{i'}\|_{\mathcal{X}} + \|x_{i'} - x_j\|_{\mathcal{X}},$$

so that $f^*(x_{i'},1) < f^*(x_j,1) + \|x_{i'} - x_j\|_{\mathcal{X}}$. But then it is possible to increase the bias by increasing $f^*(x_{i'},1)$, which cannot be the case at the optimum. If (S2) doesn't hold for some (j,j'), then by triangle inequality, for all i,

$$
\begin{aligned}
f^*(x_j,1) + \|x_i - x_j\|_{\mathcal{X}} &> f^*(x_{j'},1) + \|x_i - x_j\|_{\mathcal{X}} + \|x_j - x_{j'}\|_{\mathcal{X}}\\
&\geq f^*(x_{j'},1) + \|x_i - x_{j'}\|_{\mathcal{X}} \geq f^*(x_i,1).
\end{aligned}
$$

But this contradicts the assertion that for each j, at least one of the constraints $f(x_i,1) \leq f(x_j,1) + \|x_i - x_j\|_{\mathcal{X}}$ binds. Finally, suppose that (S3) doesn't hold for some (i,j). Then by triangle inequality,

$$
\begin{aligned}
f^*(x_i,1) + \|x_i - x_{i(j)}\|_{\mathcal{X}} &\leq f^*(x_i,1) + \|x_i - x_j\|_{\mathcal{X}} + \|x_{i(j)} - x_j\|_{\mathcal{X}}\\
&< f^*(x_j,1) + \|x_{i(j)} - x_j\|_{\mathcal{X}} = f^*(x_{i(j)},1),
\end{aligned}
$$

which violates (S1).

D.2 Proof of Lemma A.4

We will show that Equations (28), (29) and (30) hold at the optimum for $d_i, d_{i'} = 1$ and $d_j, d_{j'} = 0$. The argument that they hold for $d_i, d_{i'} = 0$ and $d_j, d_{j'} = 1$ is similar and omitted. The first-order conditions associated with the Lagrangian (31) are

$$
\begin{aligned}
&m_j/\sigma^2(0) = \mu w(0) + \sum_{i=1}^{n_1} \Lambda_{ij}^0, && \mu w(0) = \sum_{i=1}^{n_1} \Lambda_{ij}^1 && j = 1,\dots,n_0, && \text{(S4)}\\
&m_{i+n_0}/\sigma^2(1) = \mu w(1) + \sum_{j=1}^{n_0} \Lambda_{ij}^1, && \mu w(1) = \sum_{j=1}^{n_0} \Lambda_{ij}^0 && i = 1,\dots,n_1. && \text{(S5)}
\end{aligned}
$$

If $w(0) = 0$, the first-order conditions together with the dual feasibility condition $\Lambda_{ij}^1 \geq 0$ implies that $m_{i+n_0} = \mu w(1)\sigma^2(1)$, and the assertion of the lemma holds trivially, since $r_j = \mu w(1)\sigma^2(1)$

for $j = 1, \ldots, n$ achieves the optimum. Suppose, therefore, that $w(0) > 0$. Then $\sum_{i=1}^{n_1} \Lambda^1_{ij} > 0$, so that at least one of the constraints associated with Λ^1_{ij} must bind for each j. Let $i(j)$ denote the index of one of the binding constraints (picked arbitrarily if it is not unique), so that $r_j = m_{i(j)+n_0} + \|x_{i(j)+n_0} - x_j\|_{\mathcal{X}}$. Suppose (28) didn't hold, so that for some $j, j' \in \{1, \ldots, n_0\}$, $r_j > r_{j'} + \|x_j - x_{j'}\|_{\mathcal{X}}$. Then by triangle inequality

$$r_j > r_{j'} + \|x_j - x_{j'}\|_{\mathcal{X}} = m_{i(j')+n_0} + \|x_{i(j')+n_0} - x_{j'}\|_{\mathcal{X}} + \|x_j - x_{j'}\|_{\mathcal{X}} \geq m_{i(j')+n_0} + \|x_{i(j')+n_0} - x_j\|_{\mathcal{X}},$$

which violates the constraint associated with $\Lambda^1_{i(j')j}$. Next, if (29) didn't hold, so that for some $i, i' \in \{1, \ldots, n_1\}$, $m_{i+n_0} > m_{i'+n_0} + \|x_{i+n_0} - x_{i'+n_0}\|_{\mathcal{X}}$, then for all $j \in \{1, \ldots, n_0\}$,

$$r_j \leq m_{i'+n_0} + \|x_{i'+n_0} - x_j\|_{\mathcal{X}} \leq m_{i'+n_0} + \|x_{i'+n_0} - x_{i+n_0}\|_{\mathcal{X}} + \|x_{i+n_0} - x_j\|_{\mathcal{X}} < m_{i+n_0} + \|x_{i+n_0} - x_j\|_{\mathcal{X}},$$

The complementary slackness condition $\Lambda^1_{ij}(r_j - m_{i+n_0} - \|x_{i+n_0} - x_j\|_{\mathcal{X}}) = 0$ then implies that $\sum_j \Lambda^1_{ij} = 0$, and it follows from the first-order condition that $m_{i+n_0}/\sigma^2(1) = \mu w(1) \leq m_{i'+n_0}/\sigma^2(1)$, which contradicts the assertion that $m_{i+n_0} > m_{i'+n_0} + \|x_{i+n_0} - x_{i'+n_0}\|_{\mathcal{X}}$. Finally, if (30) didn't hold, so that $m_{i+n_0} > r_j + \|x_{i+n_0} - x_j\|_{\mathcal{X}}$ for some $i \in \{1, \ldots, n_1\}$ and $j \in \{1, \ldots, n_0\}$, then by triangle inequality

$$m_{i+n_0} > r_j + \|x_{i+n_0} - x_j\|_{\mathcal{X}} = m_{i(j)} + \|x_{i(j)+n_0} - x_j\|_{\mathcal{X}} + \|x_{i+n_0} - x_j\|_{\mathcal{X}} \geq m_{i(j)} + \|x_{i(j)+n_0} - x_{i+n_0}\|_{\mathcal{X}},$$

which contradicts (29).

D.3 Derivation of algorithm for solution path

Observe that $\Lambda^0_{ij} = 0$ unless for some k, $i \in \mathcal{R}^0_k$ and $j \in \mathcal{M}^0_k$, and similarly $\Lambda^1_{ij} = 0$ unless for some k, $j \in \mathcal{R}^1_k$ and $i \in \mathcal{M}^1_k$. Therefore, the first-order conditions (S4) and (S5) can equivalently be written as

$$m_j/\sigma^2(0) = \mu w(0) + \sum_{i \in \mathcal{R}^0_k} \Lambda^0_{ij} \qquad j \in \mathcal{M}^0_k, \qquad \mu w(1) = \sum_{j \in \mathcal{M}^0_k} \Lambda^0_{ij} \qquad i \in \mathcal{R}^0_k, \tag{S6}$$

$$m_{i+n_0}/\sigma^2(1) = \mu w(1) + \sum_{j \in \mathcal{R}^1_k} \Lambda^1_{ij} \qquad i \in \mathcal{M}^1_k, \qquad \mu w(0) = \sum_{i \in \mathcal{M}^1_k} \Lambda^1_{ij} \qquad j \in \mathcal{R}^1_k. \tag{S7}$$

Summing up these conditions then yields

$$\sum_{j \in \mathcal{M}^0_k} m_j/\sigma^2(0) = \mu w(0) \cdot \#\mathcal{M}^0_k + \sum_{j \in \mathcal{M}^0_k} \sum_{i \in \mathcal{R}^0_k} \Lambda^0_{ij} = \#\mathcal{M}^0_k \cdot \mu w(0) + \#\mathcal{R}^0_k \cdot \mu w(1),$$

$$\sum_{i \in \mathcal{M}^1_k} m_{i+n_0}/\sigma^2(1) = \mu w(1) \cdot \#\mathcal{M}^1_k + \sum_{i \in \mathcal{M}^1_k} \sum_{j \in \mathcal{R}^1_k} \Lambda^1_{ij} = \#\mathcal{M}^1_k \cdot \mu w(1) + \#\mathcal{R}^1_k \cdot \mu w(0).$$

Following the argument in Osborne et al. (2000, Section 4), by continuity of the solution path, for a small enough perturbation s, $N^d(\mu + s) = N^d(\mu)$, so long as the elements of $\Lambda^d(\mu)$ associated with the active constraints are strictly positive. In other words, the set of active constraints doesn't change for small enough changes in μ. Hence, the partition $\mathcal{M}_k^d$ remains the same for small enough changes in μ and the solution path is differentiable. Differentiating the preceding display yields

$$\frac{1}{\sigma^2(0)} \sum_{j\in\mathcal{M}_k^0} \frac{\partial m_j(\mu)}{\partial\mu} = \#\mathcal{M}_k^0 \cdot w(0) + \#\mathcal{R}_k^0 \cdot w(1),$$
$$\frac{1}{\sigma^2(1)} \sum_{i\in\mathcal{M}_k^1} \frac{\partial m_{i+n_0}(\mu)}{\partial\mu} = \#\mathcal{M}_k^1 \cdot w(1) + \#\mathcal{R}_k^1 \cdot w(0).$$

If $j \in \mathcal{M}_k^0$, then there exists a j' and i such that the constraints associated with Λ_{ij}^0 and $\Lambda_{ij'}^0$ are both active, so that $m_j + \|x_{i+n_0} - x_j\|_{\mathcal{X}} = r_{i+n_0} = m_{j'} + \|x_{i+n_0} - x_{j'}\|_{\mathcal{X}}$, which implies that $\partial m_j(\mu)/\partial\mu = \partial m_{j'}(\mu)/\partial\mu$. Since all elements in $\mathcal{M}_k^0$ are connected, it follows that the derivative $\partial m_j(\mu)/\partial\mu$ is the same for all j in $\mathcal{M}_k^0$. Similarly, $\partial m_j(\mu)/\partial\mu$ is the same for all j in $\mathcal{M}_k^1$. Combining these observations with the preceding display implies

$$\frac{1}{\sigma^2(0)} \frac{\partial m_j(\mu)}{\partial\mu} = w(0) + \frac{\#\mathcal{R}_{k(j)}^0}{\#\mathcal{M}_{k(j)}^0} w(1), \qquad \frac{1}{\sigma^2(1)} \frac{\partial m_{i+n_0}(\mu)}{\partial\mu} = w(1) + \frac{\#\mathcal{R}_{k(i)}^1}{\#\mathcal{M}_{k(i)}^1} w(0),$$

where $k(i)$ and $k(j)$ are the partitions that i and j belong to. Differentiating the first-order conditions (S6) and (S7) and combining them with the restriction that $\partial\Lambda_{ij}^d(\mu)/\partial\mu = 0$ if $N_{ij}^d(\mu) = 0$ then yields the following set of linear equations for $\partial\Lambda^d(\mu)/\partial\mu$:

$$\frac{\#\mathcal{R}_k^0}{\#\mathcal{M}_k^0} w(1) = \sum_{i\in\mathcal{R}_k^0} \frac{\partial\Lambda_{ij}^0(\mu)}{\partial\mu}, \qquad w(1) = \sum_{j\in\mathcal{M}_k^0} \frac{\partial\Lambda_{ij}^0(\mu)}{\partial\mu},$$
$$\frac{\#\mathcal{R}_k^1}{\#\mathcal{M}_k^1} w(0) = \sum_{j\in\mathcal{R}_k^1} \frac{\partial\Lambda_{ij}^1(\mu)}{\partial\mu}, \qquad w(0) = \sum_{i\in\mathcal{M}_k^1} \frac{\partial\Lambda_{ij}^1(\mu)}{\partial\mu}, \qquad \frac{\partial\Lambda_{ij}^d(\mu)}{\partial\mu} = 0 \qquad \text{if } N_{ij}^d(\mu) = 0.$$

Therefore, $m(\mu)$, $\Lambda^0(\mu)$, and $\Lambda^1(\mu)$ are all piecewise linear in μ. Furthermore, since for $i \in \mathcal{R}_k^0$, $r_{i+n_0}(\mu) = m_j(\mu) + \|x_{i+n_0} - x_j\|_{\mathcal{X}}$ where $j \in \mathcal{M}_k^0$, it follows that

$$\frac{\partial r_{i+n_0}(\mu)}{\partial\mu} = \frac{\partial m_j(\mu)}{\partial\mu} = \sigma^2(0)\left[w(0) + \frac{\#\mathcal{R}_k^0}{\#\mathcal{M}_k^0} w(1)\right].$$

Similarly, since for $j \in \mathcal{R}_k^1$, and $i \in \mathcal{M}_k^1$ $r_j(\mu) = m_{i+n_0}(\mu) + \|x_{i+n_0} - x_j\|_{\mathcal{X}}$, where $j \in \mathcal{M}_k^0$, we have

$$\frac{\partial r_j(\mu)}{\partial\mu} = \frac{\partial m_{i+n_0}(\mu)}{\partial\mu} = \sigma^2(1)\left[w(1) + \frac{\#\mathcal{R}_k^1}{\#\mathcal{M}_k^1} w(0)\right].$$

Thus, $r(\mu)$ is also piecewise linear in μ.

Differentiability of m and Λ^d is violated if the condition that the elements of Λ^d associated with the active constraints are all strictly positive is violated. This happens if one of the non-zero elements of $\Lambda^d(\mu)$ decreases to zero, or else if a non-active constraint becomes active, so that for some i and j with $N^0_{ij}(\mu) = 0$, $r_{i+n_0}(\mu) = m_j(\mu) + \|x_{i+n_0} - x_j\|_{\mathcal{X}}$, or for some i and j with $N^1_{ij}(\mu) = 0$, $r_j(\mu) = m_{i+n_0}(\mu) + \|x_{i+n_0} - x_j\|_{\mathcal{X}}$. This determines the step size s in the algorithm.

D.4 Proof of Lemma B.2

For ease of notation, let $f_i = f(x_i, d_i)$, $\sigma^2_i = \sigma^2(x_i, d_i)$, and let $\overline{f}_i = J^{-1}\sum_{j=1}^J f_{\ell_j(i)}$ and $\overline{u}_i = J^{-1}\sum_{j=1}^J u_{\ell_j(i)}$. Then we can decompose

$$\frac{J+1}{J}(\hat{u}^2_i - u^2_i) = [f_i - \overline{f}_i + u_i - \overline{u}_i]^2 - \frac{J+1}{J}u^2_i$$
$$= [(f_i - \overline{f}_i)^2 + 2(u_i - \overline{u}_i)(f_i - \overline{f}_i)] - 2\overline{u}_i u_i + \frac{2}{J^2}\sum_{j=1}^J\sum_{k=1}^{j-1} u_{\ell_j(i)}u_{\ell_k(i)} + \frac{1}{J^2}\sum_{j=1}^J(u^2_{\ell_j(i)} - u^2_i)$$
$$= T_{1i} + 2T_{2i} + 2T_{3i} + T_{4i} + T_{5i} + \frac{1}{J^2}\sum_{j=1}^J(\sigma^2_{\ell_j(i)} - \sigma^2_i),$$

where

$$T_{1i} = [(f_i - \overline{f}_i)^2 + 2(u_i - \overline{u}_i)(f_i - \overline{f}_i)], \qquad T_{2i} = \overline{u}_i u_i$$
$$T_{3i} = \frac{1}{J^2}\sum_{j=1}^J\sum_{k=1}^{j-1} u_{\ell_j(i)}u_{\ell_k(i)}, \qquad T_{4i} = \frac{1}{J^2}\sum_{j=1}^J(u^2_{\ell_j(i)} - \sigma^2_{\ell_j(i)}), \qquad T_{5i} = \sigma^2_i - u^2_i.$$

Since $\max_i\|x_{\ell_J(i)} - x_i\| \to 0$ and since $\sigma^2(\cdot, d)$ is uniformly continuous, it follows that

$$\max_i \max_{1\leq j\leq J}|\sigma^2_{\ell_j(i)} - \sigma^2_i| \to 0,$$

and hence that $|\sum_{i=1}^n a_{ni}J^{-1}\sum_{j=1}^J(\sigma^2_{\ell_j(i)} - \sigma^2_i)| \leq \max_i \max_{j=1,\dots,J}(\sigma^2_{\ell_j(i)} - \sigma^2_i)\sum_{i=1}^n a_{ni} \to 0$. To prove the lemma, it therefore suffices to show that the sums $\sum_{i=1}^n a_{ni}T_{qi}$ all converge to zero.

To that end,

$$E|\sum_i a_{ni}T_{1i}| \leq \max_i(f_i - \overline{f}_i)^2\sum_i a_{ni} + 2\max_i|f_i - \overline{f}_i|\sum_i a_{ni}E|u_i - \overline{u}_i|,$$

which converges to zero since $\max_i|f_i - \overline{f}_i| \leq \max_i \max_{j=1,\dots,J}(f_i - f_{\ell_j(i)}) \leq C_n \max_i\|x_i - x_{\ell_J(i)}\|_{\mathcal{X}} \to$

0. Next, by the von Bahr-Esseen inequality,

$$E|\sum_{i=1}^{n} a_{ni}T_{5i}|^{1+1/2K} \leq 2\sum_{i=1}^{n} a_{ni}^{1+1/2K} E|T_{5i}|^{1+1/2K} \leq 2\max_i a_{ni}^{1/2K} \max_j E|T_{5j}|^{1+1/2K} \sum_{k=1}^{n} a_{nk} \to 0.$$

Let $\mathcal{I}_j$ denote the set of observations for which an observation j is used as a match. To show that the remaining terms converge to zero, let we use the fact $\#\mathcal{I}_j$ is bounded by $J\overline{L}$, where $\overline{L}$ is the kissing number, defined as the maximum number of non-overlapping unit balls that can be arranged such that they each touch a common unit ball (Miller et al., 1997, Lemma 3.2.1; see also Abadie and Imbens, 2008). $\overline{L}$ is a finite constant that depends only on the dimension of the covariates (for example, $\overline{L} = 2$ if $\dim(x_i) = 1$). Now,

$$\sum_i a_{ni}T_{4i} = \frac{1}{J^2}\sum_{j=1}^{n}(u_j - \sigma_j^2)\sum_{i\in\mathcal{I}_j} a_{ni},$$

and so by the von Bahr-Esseen inequality,

$$\begin{aligned} E|\sum_i a_{ni}T_{4i}|^{1+1/2K} &\leq \frac{2}{J^{2+1/K}}\sum_{j=1}^{n} E|u_j - \sigma_j^2|^{1+1/2K}\left(\sum_{i\in\mathcal{I}_j} a_{ni}\right)^{1+1/2K} \\ &\leq \frac{(J\overline{L})^{1/2K}}{J^{2+1/K}}\max_k E|u_k - \sigma_k^2|^{1+1/2K}\max_i a_{ni}^{1+1/2K}\sum_{j=1}^{n}\sum_{i\in\mathcal{I}_j} a_{ni}, \end{aligned}$$

which is bounded by a constant times $\max_i a_{ni}^{1+1/2K}\sum_{j=1}^{n}\sum_{i\in\mathcal{I}_j} a_{ni} = \max_i a_{ni}^{1+1/2K} J\sum_i a_{ni} \to 0$. Next, since $E[u_i u_{i'} u_{\ell_j(i)} u_{\ell_k(i')}]$ is non-zero only if either $i = i'$ and $\ell_j(i) = \ell_k(i')$, or else if $i = \ell_k(i')$ and $i' = \ell_j(i)$, we have $\sum_{i'=1}^{n} a_{ni'}E[u_i u_{i'} u_{\ell_j(i)} u_{\ell_k(i')}] \leq \max_{i'} a_{ni'}\left(\sigma_i^2\sigma_{\ell_j(i)}^2 + \sigma_{\ell_j(i)}^2\sigma_i^2\right)$, so that

$$\text{var}(\sum_i a_{ni}T_{2i}) = \frac{1}{J^2}\sum_{i,j,k,i'} a_{ni}a_{ni'}E[u_i u_{\ell_k(i')} u_{i'} u_{\ell_j(i)}] \leq 2K^2\max_{i'} a_{ni'}\sum_i a_{ni} \to 0.$$

Similarly for $j \neq k$ and $j' \neq k$, $\sum_{i'=1}^{n} a_{ni'}E[u_{\ell_j(i)} u_{\ell_k(i)} u_{\ell_{j'}(i')} u_{\ell_{k'}(i')}] \leq \max_{i'} 2\sigma_{\ell_j(i)}^2\sigma_{\ell_k(i)}^2$, so that

$$\begin{aligned} &\text{var}\left(\sum_i a_{ni}T_{3i}\right) \\ &\qquad = \frac{1}{J^4}\sum_{i,i',j,j'}\sum_{k=1}^{j-1}\sum_{k'=1}^{j'-1} a_{ni}a_{ni'}E[u_{\ell_j(i)} u_{\ell_k(i)} u_{\ell_{j'}(i')} u_{\ell_{k'}(i')}] \leq 2K^2\max_{i'} a_{ni'}\sum_i a_{ni} \to 0. \end{aligned}$$

References

Abadie, A. and G. W. Imbens (2008): "Estimation of the Conditional Variance in Paired Experiments," *Annales d'Économie et de Statistique*, 175–187.

Miller, G. L., S.-H. Teng, W. Thurston, and S. A. Vavasis (1997): "Separators for Sphere-Packings and Nearest Neighbor Graphs," *Journal of the ACM*, 44, 1–29.

Osborne, M. R., B. Presnell, and B. A. Turlach (2000): "A New Approach to Variable Selection in Least Squares Problems," *IMA Journal of Numerical Analysis*, 20, 389–404.