Efficient Solution of Large Fixed Effects Problems Using R

Appendix: Matrix and Operation Density

Tom Balmat, Jerome P. Reiter

December 3, 2017

In solving the OLS normal equations, $\boldsymbol{X'X}\hat{\boldsymbol{\beta}} = \boldsymbol{X'Y}$ for $\hat{\boldsymbol{\beta}}$, using direct linear algebra operations,[1] the matrix product $\boldsymbol{X'X}$ must be computed. When $\boldsymbol{X}$ has low density (the proportion of non-zero entries is low) we are presented with an opportunity to limit computations to the small number that affect the final result, in the case of matrix multiplication, products of non-zero row and column elements. Consider a design matrix of dimension $n \times p$, such that p consists of p_x columns for continuous variables (including a constant) and p_k columns, one for each level of a single fixed effect. Note that the fixed effect actually has $p_k + 1$ levels, but one "reference" level is omitted to avoid linear dependence of the fixed effect and constant columns. Effect levels are mutually exclusive within observation since, at most, one level applies. Also, reference level observations have all fixed effect columns coded as 0. Therefore, within the np_k fixed effect cells, the maximum number of cells coded as 1 is $n - 1$ (at least one reference level observation must exist), giving a density upper bound within the fixed effect columns of $D_{M_k} \leq \frac{n-1}{np_k} < \frac{n}{np_k} = \frac{1}{p_k}$. Letting d_j be the density of the jth continuous column, overall design matrix density has an upper bound of $D_M \leq \frac{n[\sum_{j=1}^{p_x} d_j] + n - 1}{np}$. Assuming high density in the continuous columns (near 1.0), this becomes

$$D_M \leq \frac{np_x + n - 1}{np} < \frac{p_x + 1}{p} = \frac{p_x + 1}{p_x + p_k}.$$

It follows that density decreases with increasing fixed effect levels. In fact, the rate of change per additional level (assuming high density continuous columns) of the upper bound of D_M is

$$\Delta D_M = \frac{np_x + n}{n(p+1)} - \frac{np_x + n}{np} = \frac{p_x + 1}{p+1} - \frac{p_x + 1}{p} = -\frac{p_x + 1}{p(p+1)} \propto -\frac{1}{p^2} \quad (approximate)$$
$$= -\frac{1}{(p_x + p_k)^2}.$$

[1] As opposed to, for instance, employing the QR decomposition of $\boldsymbol{X}$. For more on this method, see Dept. of Statistics, University of Wisconsin (1997).

It is interesting to note that the upper bounds for both D_M and ΔD_M depend solely on p which, for high dimension problems, is dominated by p_k, making density predominantly a function of the number of fixed effect levels. The upper panel of figure 1 plots total design matrix density against number of fixed effect levels from 1 through 500 (less the reference level) for a model with 1, 2, 5, and 10 continuous variables (plus a constant term) and a single fixed effect. It is seen that density rapidly decreases as the number of levels increases, presenting efficiency opportunities for all but the lowest dimension problems. In a multiple fixed effects problem, with m effects and $p_K = \sum_{i=1}^{m} p_k i$ fixed effects columns, density is further reduced since np_K cells are populated with a maximum of $m(n-1)$ non-zero values. Given m fixed effects, an upper bound on matrix density is

$$D_M \leq \frac{np_x + m(n-1)}{n(p_x + p_k)} < \frac{p_x + m}{p_x + p_k} = \frac{p_x + m}{p}$$

The lower panel of figure 1 shows, for models with 5 continuous variables and 1, 2, 5, or 10 fixed effects, significant reduction in density as the combined number of levels, p_k, increases.

Design matrix density measures the proportion of elements that contribute information to and cannot be eliminated from the $\boldsymbol{X'X}$ operation. Figure 1 indicates that, for even low dimension fixed effects, the proportion of essential matrix elements is relatively small, typically less than 0.01 for problems in the class we consider. However, design matrix density merely characterizes an opportunity for efficient computation. More important than matrix density itself is what we will call operation density, the proportion of standard matrix operations (sums of products of row and column elements) that contribute to the resulting $\boldsymbol{X'X}$ product. Recall that, given $\boldsymbol{X}(n \times p)$, the matrix product $\boldsymbol{X'X}$ has $p \times p$ cells, the ij^{th} being the product $\boldsymbol{X}_i'\boldsymbol{X}_j$, which involves the addition of n element products, for a total of $n + n - 1$ basic operations (n multiplication and $n-1$ addition operations). The total number of basic operations to produce $\boldsymbol{X'X}$, using standard matrix methods, is then $(2n-1)p^2$. Consider a model with p_x continuous variables, each having (matrix) density of 1.0, and a single fixed effect with p_k levels that are uniformly distributed throughout observations (giving fixed effect indicator column density of $d_k = \frac{1}{p_k}$. The number of operations for products of continuous columns and continuous columns is $(2n-1)p_x^2$, for products of continuous and fixed effect indicator columns is $(2nd_k - 1)p_x p_k = (2n - p_k)p_x$, and for products of fixed effect and fixed effect columns is $(2nd_k^2 - 1)p_k^2 = 2n - p_k^2$. Operation density for this model is then

$$D_O = \frac{(2n-1)p_x^2 + (2n-p_k)p_x + 2n - p_k^2}{(2n-1)p^2}$$

$$\leq \frac{(2n-1)p_x^2 + (2n-1)p_x + 2n - 1}{(2n-1)p^2}$$

$$= \frac{p_x^2 + p_x + 1}{p^2} \propto \frac{1}{p^2}.$$

The rate of change, per additional level of the fixed effect, of the upper bound of D_O is

$$\Delta D_O = (p_x^2 + p_x + 1)\left[\frac{1}{(p+1)^2} - \frac{1}{p^2}\right]$$

$$= -(p_x^2 + p_x + 1)\frac{2p+1}{p^2(p+1)^2} \propto -\frac{1}{p^3} \quad (approximate).$$

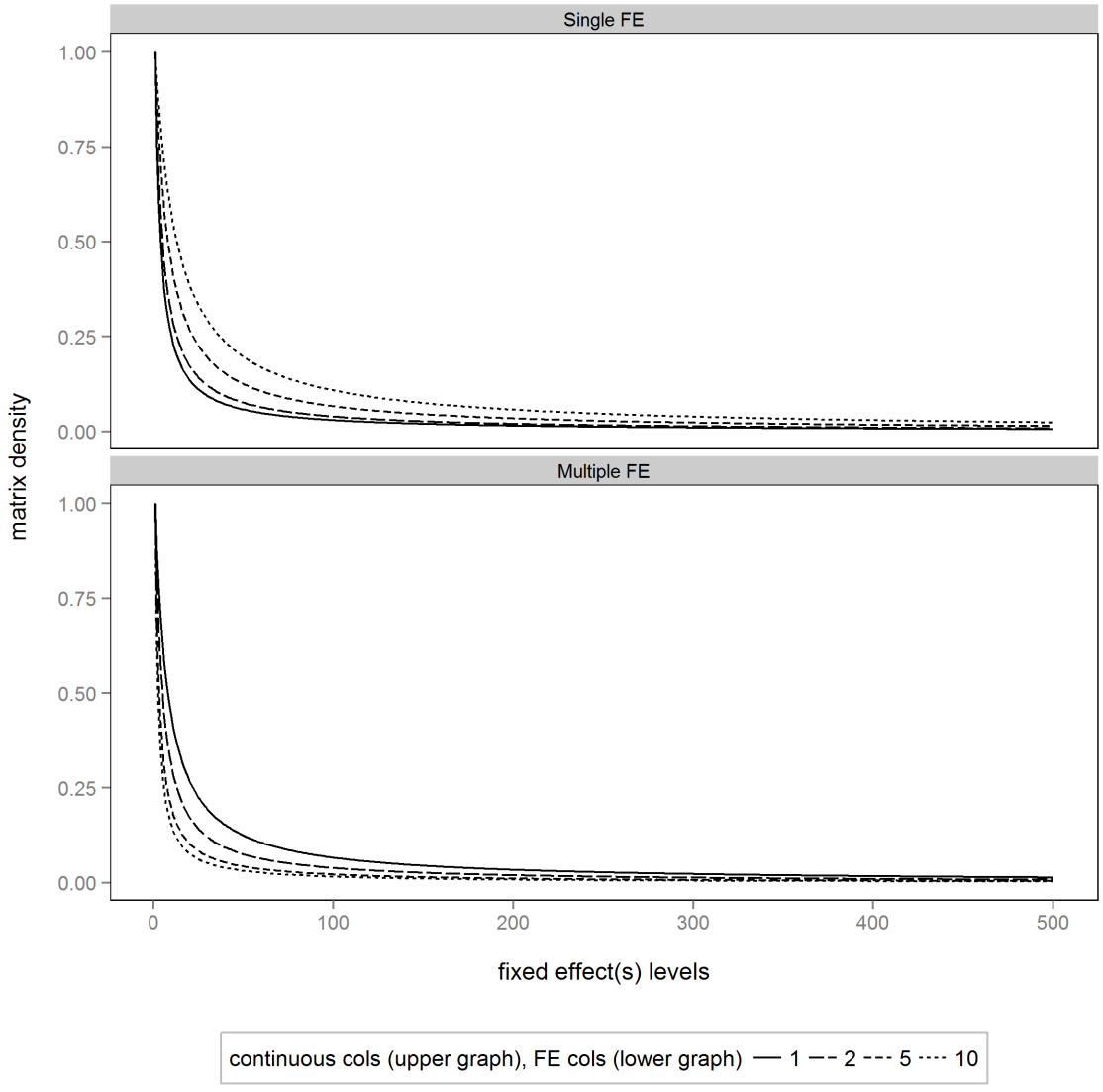

Figure 1: Design Matrix Density vs. Increase In Levels of Fixed Effects. Upper panel, one to ten continuous columns with a single fixed effect. Lower panel, a single continuous column with one to ten fixed effects with common dimension (number of levels).

Figure 2 shows the relationship of operation density to fixed effect levels. The upper panel plots ΔD_O against fixed effect levels for a model with 1, 2, 5, or 10 continuous variables and one fixed effect. The

lower panel plots ΔD_O against fixed effect levels for a model with 5 continuous variables and 1, 2, 5, or 10 fixed effects, each with dimension as indicated on the x-axis. A rapid decrease in ΔD_O is apparent even when fixed effects are of low dimension, further evidence of an opportunity of computational efficiency by operating solely on non-zero cells of $\boldsymbol{X}$. It should be noted that an increase in fixed effect levels does not increase the number of non-zero matrix elements or operations, but does increases the number of standard $\boldsymbol{X}'\boldsymbol{X}$ operations, hence the relative decrease in ΔD_O. Another important feature of ΔD_O is that since, within a given fixed effect, indicator columns are orthogonal, the p_k^2 elements of $\boldsymbol{X}'\boldsymbol{X}$ corresponding to it are necessarily diagonal ($p_k \times p_k$). This means that, for example, a model with four continuous variables (plus a constant) and one fixed effect with 96 levels has 952 95 = 8,930 of 10,000 total $\boldsymbol{X}'\boldsymbol{X}$ elements known in advance to be zero. Any operations involving elements from columns i and j within a fixed effect, where $i \neq j$, are unnecessary. Also, since a fixed effect is represented in $\boldsymbol{X}$ as a series of indicator columns, one for each level, the p_k diagonal elements of $\boldsymbol{X}'\boldsymbol{X}$ corresponding to an effect are simply $\boldsymbol{X}_j'\boldsymbol{X}_j = n_j$, where j is the indicator column index and n_j is the observation count for the corresponding fixed effect and level. Thus, multiplication operations within fixed effects can be substituted with simple counts.

We have established that matrix density is strongly influenced by fixed effect dimension, that operation density is a result of matrix density, and that theoretical measures of operation density indicate an opportunity, for sufficiently sparse problems, to eliminate a high proportion of standard matrix operations in producing $\boldsymbol{X}'\boldsymbol{X}$. Analysis of figure 2 indicates that for models involving fixed effects with 50 or more levels, the proportion of effective operations (those that must be executed) approaches zero. Models with fixed effects in this range are very common (50 or more occupations, states, counties, schools, etc.), making implementation of a solution method both relevant and appealing.

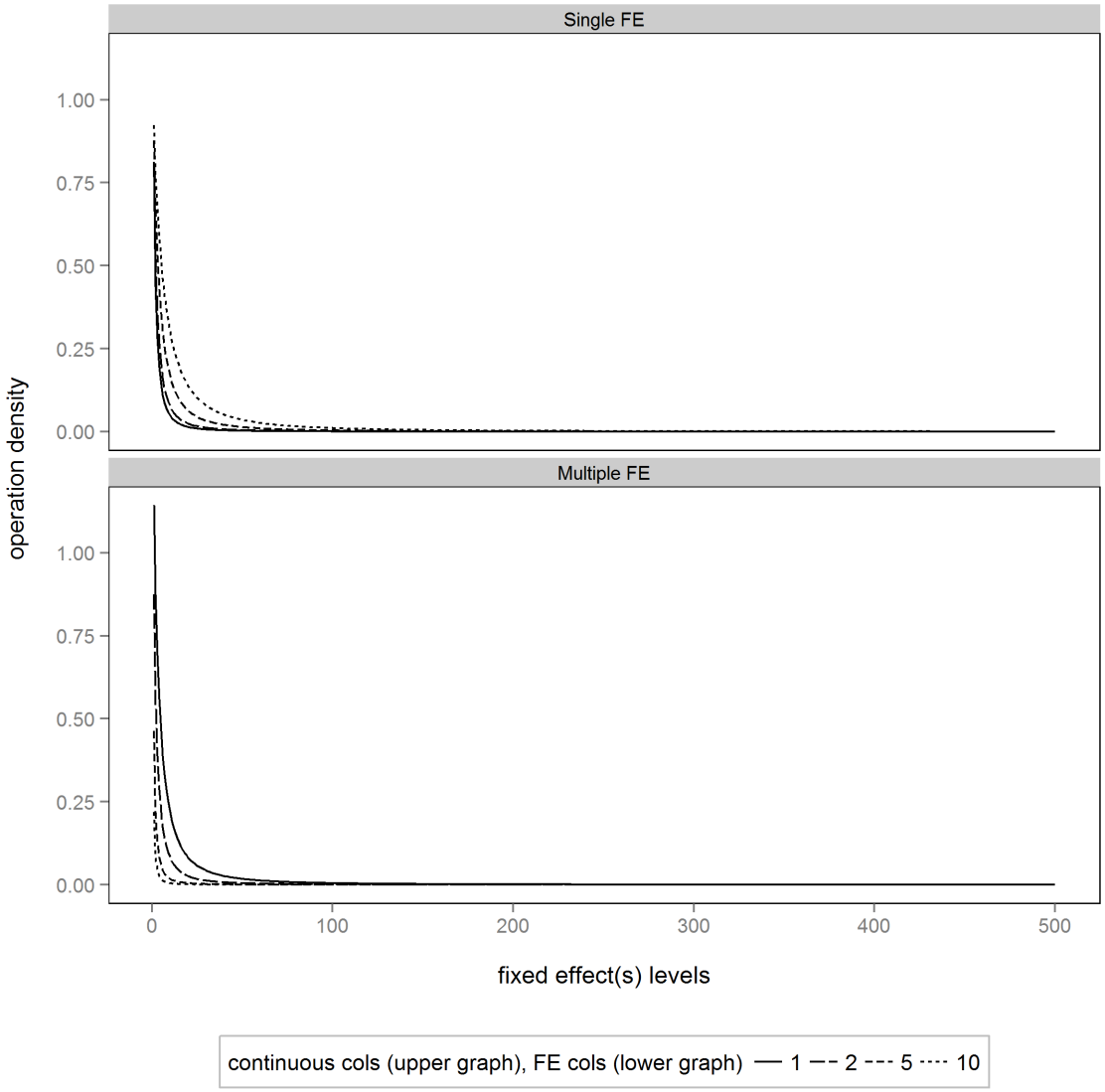

Figure 2: Operation Density vs. Increase in Fixed Effects Levels. Upper panel, one to ten continuous columns with a single fixed effect. Lower panel, a single continuous column with one to ten fixed effects with common dimension (number of levels).

References

Dept. of Statistics, University of Wisconsin. The QR Decomposition and Regression, 1997. URL `https://www.stat.wisc.edu/ larget/math496/qr.html`.