sparseMVN: An R Package for Multivariate Normal Functions with Sparse Covariance and Precision Matrices

Michael Braun
Edwin L. Cox School of Business
Southern Methodist University

Abstract

The sparseMVN package exploits sparsity in covariance and precision matrices to speed up multivariate normal simpulation and density computation.

Keywords: multivariate normal, sparse matrices.

The **mvtnorm** package (Genz, Bretz, Miwa, Mi, Leisch, Scheipl, and Hothorn 2012) provides the `dmvnorm` function to compute the density of a multivariate normal (MVN) distributon, and the `rmvnorm` function to simulate MVN random variables. These functions require the user to supply a full, "dense" covariance matrix; if the precision matrix is more readily available, the user must first invert it explicitly. This covariance matrix is dense in the sense that, for an M-dimensional MVN random variable, all M^2 elements are stored, so memory requirements grow quadratically with the size of the problem. Internally, both functions factor the covariance matrix using a Cholesky decomposition, whose complexity is $\mathcal{O}(M^3)$(Golub and Van Loan 1996).[1] This factorization is performed every time the function is called, even if it does not change from call to call. Also, `rmvnorm` involves multiplication of a triangular matrix, and `dmvnorm` involves solving a triangular linear system. Both of these operations are $\mathcal{O}(M^2)$ (Golub and Van Loan 1996). MVN functions in other packages, such as **MASS** (Venables and Ripley 2002) and **LaplacesDemon** (Statisticat and LLC. 2016), face similar limitations.[2] Thus, existing tools for working with the MVN distribution in R are not practical for high-dimensional MVN random variables.

However, for many applications the covariance or precision matrix is sparse, meaning that the proportion of nonzero elements is small, relative to the total size of the matrix. The **sparseMVN** package exploits that sparsity to reduce memory requirements, and to gain computational efficiencies, when computing the MVN density (`dmvn.sparse`), and simulating from an MVN random variable (`rmvn.sparse`). Instead of requiring the user to supply a dense covariance matrix, `dmvn.sparse` and `rmvn.sparse` accept a pre-computed Cholesky decomposition of either the covariance or precision matrix in a compressed sparse format. This approach has several advantages:

[1] `dmvnorm` has options for eigen and singular value decompositions. These are both $\mathcal{O}(M^3)$ as well.

[2] **LaplacesDemon** does offer options for the user to supply pre-factored covariance and precision matrices. This avoids repeated calls to the $\mathcal{O}(M^3)$ factorization step, but not the $\mathcal{O}(M^2)$ matrix multiplication and linear system solution steps.

1. Memory requirements are lower because only the nonzero elements of the matrix are stored in a compressed sparse format.
2. Linear algebra algorithms that are optimzed for sparse matrices are more efficient because they avoid operations on matrix elements that are known to be zero.
3. When the precision matrix is initially available, there is no need to invert it into a covariance matrix explicitly. This feature of **sparseMVN** preserves sparsity, because the inverse of a sparse matrix is not necessarily sparse.
4. The Cholesky factor of the matrix is computed once, before the first **sparseMVN** function call, and is not repeated with any subsequent calls (as long as the matrix does not change).

The functions in **sparseMVN** rely on sparse matrix classes and functions defined in the **Matrix** package (Bates and Maechler 2015). The user creates the covariance or precision matrix as a sparse, symmetric *dsCMatrix* matrix, and computes the sparse Cholesky factor using the `Cholesky` function. Other than ensuring that the factor for the covariance or precision matrix is in the correct format, the `dmvn.sparse` and `rmvn.sparse` functions behave in much the same way as the corresponding **mvtnorm** functions `dmvnorm` and `rmvnorm`. Internally, **sparseMVN** uses standard methods of computing the MVN density and simulating MVN random variables (see Section XX), except that sparse-optimized algorithms are used for linear algebra operations.

1. Background

Let $x \in \mathbb{R}^M$ be a realization of random variable $X \sim \mathbf{MVN}(\mu, \Sigma)$, where $\mu \in \mathbb{R}^M$ is a vector, $\Sigma \in \mathbb{R}^{M \times M}$ is a positive-definite covariance matrix, and $\Sigma^{-1} \in \mathbb{R}^{M \times M}$ is a positive-definite precision matrix.

The log probability density of x is

$$\log f(x) = -\frac{1}{2}\left(M \log(2\pi) + \log|\Sigma| + z^\top z\right), \quad \text{where } z^\top z = (x-\mu)^\top \Sigma^{-1}(x-\mu) \tag{1}$$

1.1. MVN density computation and random number generation

The two computationally intensive steps in evaluating $\log f(x)$ are computing $\log|\Sigma|$, and $z^\top z$, *without* explicitly inverting Σ or repeating mathematical operations. How one performs these steps *efficiently* in practice depends on whether the covariance matrix Σ, or the precision matrix Σ^{-1} is available. For both cases, we start by finding a lower triangular matrix root: $\Sigma = LL^\top$ or $\Sigma^{-1} = \Lambda\Lambda^\top$. Since Σ and Σ^{-1} are positive definite, we will use the Cholesky decomposition, which is the unique matrix root with all positive elements on the diagonal.

With the Cholesky decomposition in hand, we can then compute the log determinant of Σ by adding the logs of the diagonal elements of the factors.

$$\log|\Sigma| = \begin{cases} 2\sum_{m=1}^{M} \log L_{mm} & \text{when } \Sigma \text{ is given} \\ -2\sum_{m=1}^{M} \log \Lambda_{mm} & \text{when } \Sigma^{-1} \text{ is given} \end{cases} \tag{2}$$

Having already computed the triangular matrix roots also speeds up the computation of $z^\top z$. If Σ^{-1} is given, $z = \Lambda^\top(x - \mu)$ can be computed efficiently as the product of an upper triangular matrix and a vector. When Σ is given, we find z by solving the lower triangular system $Lz = x - \mu$. The subsequent $z^\top z$ computation is trivially fast.

The algorithm for simulating $X \sim \mathbf{MVN}(\mu, \Sigma)$ also depends on whether Σ or Σ^{-1} is given. As above, we start by computing the Cholesky decomposition of the given covariance or precision matrix. Define a random variable $Z \sim \mathbf{MVN}(0, I_M)$, and generate a realization z as a vector of M samples from a standard normal distribution. If Σ is given, then evaluate $x = Lz + \mu$. If Σ^{-1} is given, then solve for x in the triangular linear system $\Lambda^\top(x - \mu) = z$. The resulting x is a sample from $\mathbf{MVN}(\mu, \Sigma)$. We confirm the mean and covariance of X as follows:

$$\mathbf{E}(X) = \mathbf{E}(LZ + \mu) = \mathbf{E}\Big(\Lambda^\top Z + \mu\Big) = \mu \tag{3}$$

$$\mathbf{cov}(X) = \mathbf{cov}(LZ + \mu) = \mathbf{E}\Big(LZZ^\top L^\top\Big) = LL^\top = \Sigma \tag{4}$$

$$\mathbf{cov}(X) = \mathbf{cov}\Big(\Lambda^{\top^{-1}} Z + \mu\Big) = \mathbf{E}\Big(\Lambda^{\top^{-1}} ZZ^\top \Lambda^{-1}\Big) = \Lambda^{\top^{-1}}\Lambda^{-1} = (\Lambda\Lambda^\top)^{-1} = \Sigma \tag{5}$$

These algorithms apply when the covariance/precision matrix is either sparse or dense. When the matrix is dense, the computational complexity is $\mathcal{O}\big(M^3\big)$ for a Cholesky decomposition, and $\mathcal{O}\big(M^2\big)$ for either solving the triangular linear system or multiplying a triangular matrix by another matrix (Golub and Van Loan 1996). Thus, the computational cost grows cubically with M before the decomposition step, and quadratically if the decomposition has already been completed. Additionally, the storage requirement for Σ (or Σ^{-1}) grows quadratically with M.

1.2. Sparse matrices in R

The **Matrix** package (Bates and Maechler 2015) defines various classes for storing sparse matrices in compressed formats. The most important one for our purposes is a *dsCMatrix*, which defines a symmetric matrix, with numeric (double precision) elements, in a column-compressed format. Three vectors define the underlying matrix: the unique nonzero values (just one triangle is needed), the indices in the value vector for the first value in each column, and the indices of the rows in which each value is located. Roughly speaking, the storage requirements for a sparse $M \times M$ symmetric matrix with V unique nonzero elements in one triangle are for V double precision numbers, $V + M + 1$ integers, and some metadata. In contrast, a dense representation of the same matrix stores M^2 double precision values, regardless of symmetry and the number of zeros. If V grows more slowly than M^2, the matrix becomes increasingly sparse (a smaller percentage of elements are nonzero), and there are greater efficiency gains from storing the matrix in a compressed sparse format.

An example

To illustrate how sparse matrices require less memory resources when compressed than when stored densely, consider the following example, which is borrowed heavily from the vignette of the **sparseHessianFD** package (Braun 2016).

Suppose we have a dataset of N households, each with T opportunities to purchase a particular product. Let y_i be the number of times household i purchases the product, out of

the T purchase opportunities, and let p_i be the probability of purchase. The heterogeneous parameter p_i is the same for all T opportunities, so y_i is a binomial random variable.

Let $\beta_i \in \mathbb{R}^k$ be a heterogeneous coefficient vector that is specific to household i, such that $\beta_i = (\beta_{i1}, \ldots, \beta_{ik})$. Similarly, $w_i \in \mathbb{R}^k$ is a vector of household-specific covariates. Define each p_i such that the log odds of p_i is a linear function of β_i and w_i, but does not depend directly on β_j and w_j for another household $j \neq i$.

$$p_i = \frac{\exp(w_i'\beta_i)}{1 + \exp(w_i'\beta_i)}, \; i = 1...N \tag{6}$$

The coefficient vectors β_i are distributed across the population of households following a multivariate normal distribution with mean $\mu \in \mathbb{R}^k$ and covariance $\mathbf{A} \in \mathbb{R}^{k \times k}$. Assume that we know $\mathbf{A}$, but not μ, so we place a multivariate normal prior on μ, with mean 0 and covariance $\mathbf{\Omega} \in \mathbb{R}^{k \times k}$. Thus, the parameter vector $x \in \mathbb{R}^{(N+1)k}$ consists of the Nk elements in the N β_i vectors, and the k elements in μ.

The log posterior density, ignoring any normalization constants, is

$$\log \pi(\beta_{1:N}, \mu | \mathbf{Y}, \mathbf{W}, \mathbf{A}, \mathbf{\Omega}) = \sum_{i=1}^{N} \left(p_i^{y_i} (1 - p_i)^{T - y_i} - \frac{1}{2} (\beta_i - \mu)^\top \mathbf{A}^{-1} (\beta_i - \mu) \right) - \frac{1}{2} \mu^\top \mathbf{\Omega}^{-1} \mu \tag{7}$$

Because one element of β_i can be correlated with another element of β_i (for the same unit), we allow for the cross-partials between elements of β_i for any i to be nonzero. Also, because the mean of each β_i depends on μ, the cross-partials between μ and any β_i can be nonzero. However, since the β_i and β_j are independent samples, and the y_i are conditionally independent, the cross-partial derivatives between an element of β_i and any element of any β_j for $j \neq i$, must be zero. When N is much greater than k, there will be many more zero cross-partial derivatives than nonzero, and the Hessian of the log posterior density will be sparse.

The sparsity pattern depends on how the variables are ordered. One such ordering is to group all of the coefficients in the β_i for each unit together, and place μ at the end.

$$\beta_{11}, \ldots, \beta_{1k}, \beta_{21}, \ldots, \beta_{2k}, \;\; \ldots \;\; , \beta_{N1}, \ldots, \beta_{Nk}, \mu_1, \ldots, \mu_k \tag{8}$$

In this case, the Hessian has a "block-arrow" structure. For example, if $N = 5$ and $k = 2$, then there are 12 total variables, and the Hessian will have the "block-arrow" pattern in Figure 1a.

Another possibility is to group coefficients for each covariate together.

$$\beta_{11}, \ldots, \beta_{N1}, \beta_{12}, \ldots, \beta_{N2}, \;\; \ldots \;\; , \beta_{1k}, \ldots, \beta_{Nk}, \mu_1, \ldots, \mu_k \tag{9}$$

Now the Hessian has an "banded" sparsity pattern, as in Figure 1b.

In both cases, the number of nonzeros is the same. There are 144 elements in this symmetric matrix. If the matrix is stored in the standard **base** R dense format, memory is reserved for all 144 values, even though only 64 values are nonzero, and only 38 values are unique. For larger matrices, the reduction in memory requirements by storing the matrix in a sparse format can be substantial.[3]. If $N = 1,000$, then $M = 2,002$, with more than 4 million elements in the

[3]Because sparse matrix structures store row and column indices of the nonzero values, they may use more memory than dense storage if the total number of elements is small

```
#
#  [1,] | | . . . . . . . . . | |
#  [2,] | | . . . . . . . . . | |
#  [3,] . . | | . . . . . . . | |
#  [4,] . . | | . . . . . . . | |
#  [5,] . . . . | | . . . . . | |
#  [6,] . . . . | | . . . . . | |
#  [7,] . . . . . . | | . . . | |
#  [8,] . . . . . . | | . . . | |
#  [9,] . . . . . . . . | | | | |
# [10,] . . . . . . . . | | | | |
# [11,] | | | | | | | | | | | |
# [12,] | | | | | | | | | | | |
```

(a) A "block-arrow" sparsity pattern.

```
#
#  [1,] | . . . . | . . . . | |
#  [2,] . | . . . . | . . . | |
#  [3,] . . | . . . . | . . | |
#  [4,] . . . | . . . . | . | |
#  [5,] . . . . | . . . . | | |
#  [6,] | . . . . | . . . . | |
#  [7,] . | . . . . | . . . | |
#  [8,] . . | . . . . | . . | |
#  [9,] . . . | . . . . | . | |
# [10,] . . . . | . . . . | | |
# [11,] | | | | | | | | | | |
# [12,] | | | | | | | | | | |
```

(b) A "banded" sparsity pattern.

Figure 1: Two examples of sparsity patterns for a hierarchical model.

Hessian. However, only 12,004 of those elements are nonzero, with 7,003 unique values in the lower triangle. The dense matrix requires 30.6 Mb of RAM, while a sparse symmetric matrix of the *dsCMatrix* class requires only 91.5 Kb.

This example is relevant because, when evaluated at the posterior mode, the Hessian matrix of the log posterior is the precision matrix of a MVN approximatation to the posterior distribution of $(\beta_{1:N}, \mu)$. If we were to simulate from this MVN using the **mvtnorm** function `rmvnorm`, or evaluate MVN densities using `dmvnorm`, we would first need to invert the dense Hessian to get the covariance matrix Σ. Internally, these functions invoke dense linear algebra routines, including matrix factorization.

2. Using the sparseMVN package

The `rmvn.sparse` generates random simulates for an MVN distribution, and `dmvn.sparse` computes the MVN log density. The signatures are

```
rmvn.sparse(n, mu, CH, prec=TRUE)
dmvn.sparse(x, mu, CH, prec=TRUE, log=TRUE)
```

The `rmvn.sparse` function returns a matrix x with `n` rows and `length(mu)` columns. `dmvn.sparse` returns a vector of length `n`: densities if `log=FALSE`, and log densities if `log=TRUE`.

The arguments are summarized in Table 1. These functions do require the user to compute the Cholesky decomposition beforehand, but this needs to be done only once (as long as Σ or Σ^{-1} does not change). `CH` should be computed using the `Cholesky` function from the **Matrix** package. More details about the `Cholesky` function are available in the **Matrix** documentation, but it is a simple function to use. The first argument is a sparse symmetric Matrix stored as a *dsCMatrix* object. As far as we know, there is no particular need to deviate from the defaults of the remaining arguments. If `Cholesky` uses a fill-reducing permutation

x	A numeric matrix. Each row is an MVN sample.
mu	A numeric vector. The mean of the MVN random variable.
CH	Either a *dCHMsimpl* or *dCHMsuper* object representing the Cholesky decomposition of the covariance/precision matrix.
prec	Logical value that identifies CH as the Cholesky decomposition of either a covariance (Σ, `prec=TRUE`) or precision(Σ^{-1}, `prec=FALSE`) matrix.
n	Number of random samples to be generated.
log	If `log=TRUE`, the log density is returned.

Table 1: Arguments to the `rmvn.sparse` and `dmvn.sparse` functions.

to compute `CH`, the functions in **sparseMVN** will handle that directly, with no additional user intervention required. The `chol` function in **base** R should not be used.

2.1. An example

Suppose we want to generate samples from an MVN approximation to the posterior distribution of our example model. The package includes functions to simulate data for the example (`binary.sim`, and to compute the log posterior density (`binary.f`), gradient (`binary.grad`), and Hessian (`binary.hess`). The `trust.optim` function in the **trustOptim** package (Braun 2014) is a nonlinear optimizer that estimates the curvature of the objective function using a sparse Hessian.

```
R> D <- binary.sim(N=50, k=2, T=50)
R> priors <- list(inv.Sigma=diag(2), inv.Omega=diag(2))
R> start <- rep(c(-1,1),51)
R> opt <- trust.optim(start,
+                     fn=sparseMVN::binary.f,
+                     gr=sparseMVN::binary.grad,
+                     hs=sparseMVN::binary.hess,
+                     data=D, priors=list(inv.Sigma=diag(2),
+                                         inv.Omega=diag(2)),
+                     method="Sparse",
+                     control=list(function.scale.factor=-1))
```

The posterior mode, and the Hessian evaluated at that point, are returned by `trust.optim`. They serve as the mean and the negative precision of the MVN approximation to the posterior distribution of the model.

```
R> R <- 100
R> pm <- opt[["solution"]]
R> H <- -opt[["hessian"]]
R> CH <- Cholesky(H)
R> samples <- rmvn.sparse(R, pm, CH, prec=TRUE)
```

We can then compute the MVN log density for each sample.

```
R> logf <- dmvn.sparse(samples, pm, CH, prec=TRUE)
```

The ability to accept a precision matrix, rather than having to invert it to a covariance matrix, is a valuable feature of **sparseMVN**. This is because the inverse of a sparse matrix is not necessarily sparse. In the following chunk, we invert the Hessian, and drop zero values to maintain any remaining sparseness. Note that there are 10,404 total elements in the Hessian.

```
R> Matrix::nnzero(H)

# [1] 402

R> Hinv <- drop0(solve(H))
R> Matrix::nnzero(Hinv)

# [1] 10404
```

Nevertheless, we should check that the log densities from `dmvn.sparse` correspond to those that we would get from `dmvnorm`.

```
R> logf_dense <- dmvnorm(samples, pm, as.matrix(Hinv), log=TRUE)
R> all.equal(logf, logf_dense)

# [1] TRUE
```

3. Timing

In this section we show the efficiency gains from **sparseMVN** by comparing the run times between `rmvn.sparse` and `rmvnorm`, and between `dmvn.sparse` and `dmvnorm`. In these tests, we construct covariance and precision matrices with a block-arrow structure, as in Figure 1a. Each block is $k \times k$, with k rows on the bottom and right margins. Cases vary in the number and size of blocks, and whether the matrix is a covariance or a precision. Table 2 summarizes the case conditions, along with the dimension of the random variable ($M = Nk + k$), and number of nonzero elements in the matrix.

Table 3 presents run times (milliseconds) to compute 1,000 MVN densities, and generate 1,000 MVN samples, using sparse (`rmvn.sparse`, `dmvn.sparse`) and dense (`rmvnorm`, `dmvnorm`) methods. For times in Table 3a, the covariance matrix was provided to the sparse functions; Table 3b times started with the precision matrix.

For cases with smaller matrices, the dense routines ran faster than the sparse ones. This is because the matrices themselves are still somewhat dense, and the sparse linear algebra routines are not optimized for them. However, as the number of blocks increases, the matrices become increasingly sparse, and the `dmvn.sparse` and `rmvn.sparse` functions run faster than their dense counterparts.

Cholesky decompositions (both dense and sparse) and inverting matrices (dense only) are the most computationally intensive steps of random number generation and density computation. In Table 4, we show how these times vary with the sparsity of the matrix.

N	k	M	M^2	nnz	nnz (LT)	% nnz
25	2	52	2,704	304	178	0.112
100	2	202	40,804	1,204	703	0.030
250	2	502	252,004	3,004	1,753	0.012
500	2	1,002	1,004,004	6,004	3,503	0.006
1,000	2	2,002	4,008,004	12,004	7,003	0.003
25	4	104	10,816	1,216	660	0.112
100	4	404	163,216	4,816	2,610	0.030
250	4	1,004	1,008,016	12,016	6,510	0.012
500	4	2,004	4,016,016	24,016	13,010	0.006
1,000	4	4,004	16,032,016	48,016	26,010	0.003

Table 2: Cases for timing comparision. N is the number of blocks in the block-arrow structure (analogous to heterogeneous units in the binary choice example), k is the size of each block. The total number of variables is $M = Nk + k$, and M^2 is the total number of elements in the matrix. nnz and nnz (LT) are the numbers of nonzero elements in the full matrix, and lower triangle, respectively. % nnz is the proportion of elements that are nonzero.

Code to replicate Tables 3 and 4 is available as an online supplement to this paper, and in the `doc/` directory of the installed package.

4. Other packages for creating and using sparse matrices

4.1. sparseHessianFD

Suppose you have a objective function that has a sparse Hessian (e.g., the log posterior density for a hierarchical model). You have an R function that computes the value of the objective, and another function that computes its gradient. You may also need the Hessian, either for a nonlinear optimization routine, or as the negative precision matrix of an MVN approximation.

It's hard enough to get the gradient, but the derivation of the Hessian might be too tedious or complicated for it to be worthwhile. However, it should not be too hard to identify which elements of the Hessian are nonzero. If you have both the gradient, and the Hessian emphpattern, then you can use the **sparseHessianFD** package (Braun 2016) to estimate the Hessian itself.

The **sparseHessianFD** package estimates the Hessian numerically, but in a way that exploits the fact that the Hessian is sparse, and that the pattern is known. The package contains functions that return the Hessian as a sparse *dgCMatrix*. This object can be coerced into a *dsCMatrix*, which in turn can be used by `rmvn.sparse` and `dmvn.sparse`.

4.2. trustOptim

The **trustOptim** package provides a nonlinear optimization routine that takes the Hessian as a sparse *dgCMatrix* object. This optimizer is useful for unconstrained optimization of a high-dimensional objective function with a sparse Hessian. It uses a trust region algorithm,

				compute density				random sample			
				dense		sparse		dense		sparse	
N	k	nvars	nnz	**mean**	sd	**mean**	sd	**mean**	sd	**mean**	sd
25	2	52	304	**15**	12	**29**	56	**58**	64	**71**	94
100	2	202	1,204	**76**	108	**102**	150	**257**	167	**249**	172
250	2	502	3,004	**338**	241	**185**	177	**802**	280	**518**	232
500	2	1,002	6,004	**831**	336	**226**	157	**1,653**	406	**686**	258
1,000	2	2,002	12,004	**1,787**	560	**245**	148	**3,608**	911	**921**	368
25	4	104	1,216	**55**	109	**54**	95	**132**	125	**151**	143
100	4	404	4,816	**237**	243	**159**	184	**581**	273	**493**	299
250	4	1,004	12,016	**740**	258	**236**	180	**1,558**	294	**760**	277
500	4	2,004	24,016	**1,919**	569	**274**	162	**3,766**	937	**960**	369
1,000	4	4,004	48,016	**2,967**	1,426	**232**	132	**7,153**	3,364	**863**	457

(a) Starting with covariance matrix.

				compute density				random sample			
				dense		sparse		dense		sparse	
N	k	nvars	nnz	**mean**	sd	**mean**	sd	**mean**	sd	**mean**	sd
25	2	52	304	**29**	87	**40**	97	**59**	67	**66**	82
100	2	202	1,204	**69**	111	**110**	152	**225**	130	**257**	182
250	2	502	3,004	**239**	194	**222**	207	**628**	243	**560**	260
500	2	1,002	6,004	**629**	275	**298**	218	**1,478**	316	**771**	274
1,000	2	2,002	12,004	**1,442**	470	**275**	166	**3,413**	749	**928**	365
25	4	104	1,216	**45**	93	**66**	111	**127**	134	**182**	184
100	4	404	4,816	**162**	172	**173**	163	**533**	243	**468**	218
250	4	1,004	12,016	**725**	290	**307**	221	**1,643**	459	**850**	306
500	4	2,004	24,016	**1,444**	430	**286**	154	**3,335**	667	**967**	300
1,000	4	4,004	48,016	**2,348**	1,058	**243**	141	**6,797**	3,271	**751**	350

(b) Starting with precision matrix.

Table 3: Time (milliseconds) to compute 1,000 MVN densities, and simulate 1,000 MVN samples, using sparse (`rmvn.sparse`, `dmvn.sparse`) or dense (`rmvnorm`, `dmvnorm`) methods, for a covariance or precision matrix with a block-arrow structure. N is number of blocks (heterogeneous units), and k is the size of each block (heterogeneous variables). nvars and nnz are, respectively, the size of the random vector, and number of nonzero elements in the covariance/precision matrix. Means and standard deviations are across XX replications.

				Cholesky				invert			
				dense		sparse		dense		sparse	
N	k	nvars	nnz.pct	**mean**	sd	**mean**	sd	**mean**	sd	**mean**	sd
25	2	52	0.112	**0.4**	3.1	**0.3**	2.6	**21.8**	56.2	**4.5**	8.1
100	2	202	0.030	**19.9**	52.7	**0.3**	2.6	**63.9**	73.8	**11.4**	40.7
250	2	502	0.012	**42.5**	56.1	**4.1**	41.5	**178.4**	123.3	**30.6**	50.4
500	2	1,002	0.006	**107.4**	93.5	**1.0**	4.3	**579.1**	219.7	**98.0**	106.7
1,000	2	2,002	0.003	**346.8**	185.3	**0.9**	2.8	**1,794.3**	427.6	**228.9**	112.9
25	4	104	0.112	**0.8**	3.9	**0.4**	2.3	**28.2**	35.6	**11.2**	40.4
100	4	404	0.030	**27.7**	25.0	**8.9**	79.5	**128.8**	126.7	**30.5**	57.2
250	4	1,004	0.012	**116.6**	108.5	**1.2**	4.4	**531.5**	204.3	**102.6**	105.1
500	4	2,004	0.006	**381.1**	223.9	**1.2**	3.8	**1,825.3**	427.4	**283.7**	167.9
1,000	4	4,004	0.003	**695.0**	372.0	**0.9**	2.2	**5,373.0**	2,500.3	**448.6**	262.1

Table 4: Time (milliseconds) for a Cholesky decompositon and matrix inversion of a sparse (*dsCMatrix*) or dense matrix. Inversion was computed with the `solve` function.

which may be more stable and robust than line search approaches. Also, it applies a stopping rule based on the norm of the gradient, as opposed to whether the algorithm makes "sufficient progress." (Many optimizers, especially `optim` in **base** R, stop too early, before the gradient is truly flat).

5. Other

Since a large proportion of elements in the matrix are zero, we need to store only the row and column indices, and the values, of the unique nonzero elements. The efficiency gains in **sparseMVN** come from storing the covariance or precision matrix in a compressed format without explicit zeros, and applying linear algebra routines that are optimized for those sparse matrix structures. The **Matrix** package calls sparse linear algebra routines that are implemented in the **CHOLMOD** library (Chen, Davis, Hager, and Rajamanickam 2008; Davis and Hager 1999, 2009); more information about these routines is available there.

The exact amount of time and memory that are saved by saving the covariance/precision matrix in a sparse format depends on the sparsity pattern. But for the hierarchical model example from earlier in this section, the number of nonzero elements grows only linearly with N. The result is that all of the steps of sampling from an MVN also grow linearly with N. Section 4 of Braun and Damien (2016) explains why this is so.

1. Each call to the function involves a new matrix factorization step. This can be costly for applications in which x or μ changes from call to call, but Σ does not.

2. In some applications, the precision matrix Σ^{-1}, and not the covariance matrix Σ, is readily available (e.g., estimating the asymptotic covariance from the inverse of a Hessian of a maximum likelihood estimator). To use the **mvtnorm** functions, Σ^{-1} would first have to be inverted explicitly.

3. The **mvtnorm** functions treat Σ as if it were dense, even if there is a large proportion of structural zeros.

The **sparseMVN** package addresses these limitations.

1. The `rmvn.sparse` and `dmvn.sparse` functions take as their matrix argument a sparse Cholesky decomposition. The user does need to do this explicitly beforehand, but once it is done, it does not have to be done again.

2. Both functions include an argument to identify the sparse Cholesky decomposition as a factor of a covariance matrix (`prec=FALSE`) or precision matrix (`prec=TRUE`).

Even when either Σ or Σ^{-1} are dense, there may be advantages to using **sparseMVN** instead of **mvtnorm**. For example, if the user were starting with a large, dense precision matrix Σ^{-1}, and computing MVN densities repeatedly, it may take less time to compute the Cholesky decomposition of $\Sigma^{-1]}$ once, than to invert it and have the **mvtnorm** functions decompose Σ over and over. Nevertheless, the main purpose of **sparseMVN** is to exploit sparsity in either Σ or Σ^{-1} when it exists.

References

Bates D, Maechler M (2015). ***Matrix****: Sparse and Dense Matrix Classes and Methods*. R package version 12-4, URL https://CRAN.R-project.org/package=Matrix.

Braun M (2014). "trustOptim: An R Package for Trust Region Optimization with Sparse Hessians." *Journal of Statistical Software*, **60**(4), 1–16. URL http://www.jstatsoft.org/v60/i04/.

Braun M (2016). *sparseHessianFD: An R package for estimating sparse Hessians*. URL https://cran.r-project.org/package=sparseHessianFD.

Braun M, Damien P (2016). "Scalable Rejection Sampling for Bayesian Hierarchical Models." *Marketing Science*, **35**(3), 427–444. doi:10.1287/mksc.2014.0901.

Chen Y, Davis TA, Hager WW, Rajamanickam S (2008). "Algorithm 887: CHOLMOD, Supernodal Sparse Cholesky Factorization and Update/Downdate." *ACM Transactions on Mathematical Software*, **35**(3), 1–14. doi:10.1145/1391989.1391995.

Davis TA, Hager WW (1999). "Modifying a Sparse Cholesky Factorization." *SIAM Journal on Matrix Analysis and Applications*, **20**(3), 606–627. doi:10.1137/S0895479897321076.

Davis TA, Hager WW (2009). "Dynamic Supernodes in Sparse Cholesky Update/Downdate and Triangular Solves." *ACM Transactions on Mathematical Software*, **35**(4), 1–23. doi:10.1145/1462173.1462176.

Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, Hothorn T (2012). *mvtnorm: Multivariate Normal and t Distributions*.

Golub GH, Van Loan CF (1996). *Matrix Computations*. 3rd edition. Johns Hopkins University Press.

Statisticat, LLC (2016). *LaplacesDemon: Complete Environment for Bayesian Inference*. R package version 16.0.1, URL https://cran.r-project.org/package=LaplacesDemon.

Venables WN, Ripley BD (2002). *Modern Applied Statistics with S*. Fourth edition. Springer-Verlag.

Affiliation:

Michael Braun
Edwin L. Cox School of Business
Southern Methodist University
6212 Bishop Blvd.
Dallas, TX 75275
E-mail: braunm@smu.edu
URL: http://www.smu.edu/Cox/Departments/FacultyDirectory/BraunMichael