

# 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 simulation 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) distribution, and the `rmvnorm` function to simulate MVN random variables. These functions require the user to supply a full, “dense” covariance matrix; if starting with a precision matrix, 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).<sup>1</sup> This factorization is performed every time the function is called, even if the covariance matrix 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) on dense matrices. MVN functions in other packages, such as **MASS** (Venables and Ripley 2002) and **LaplacesDemon** (Statisticat and LLC. 2016), face similar limitations.<sup>2</sup> 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 functions in the **sparseMVN** package exploit that sparsity to reduce memory requirements, and to gain computational efficiencies. The `dmvn.sparse` function computes the MVN density, and the `rmvn.sparse` function samples from an MVN random variable. 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:

---

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

<sup>2</sup>**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 smaller because only the nonzero elements of the matrix are stored in a compressed sparse format.
2. Linear algebra algorithms that are optimized 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 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 **sparseMVN** functions behave in much the same way as the corresponding **mvtnorm** functions. Internally, **sparseMVN** uses standard methods of computing the MVN density and simulating MVN random variables (see Section 1.1), except that sparse-optimized algorithms from the **Matrix** package 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

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

### 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  $\Sigma$  or repeating mathematical operations. How one performs these steps *efficiently* in practice depends on whether the covariance matrix  $\Sigma$ , or the precision matrix  $\Sigma^{-1}$  is available. For both cases, we start by finding a lower triangular matrix root:  $\Sigma = LL^\top$  or  $\Sigma^{-1} = \Lambda\Lambda^\top$ . Since  $\Sigma$  and  $\Sigma^{-1}$  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 compute the log determinant of  $\Sigma$  by adding the logs of the diagonal elements of the factors.

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

Having already computed the triangular matrix roots also speeds up the computation of  $z^\top z$ . If  $\Sigma^{-1}$  is given,  $z = \Lambda^\top(x - \mu)$  can be computed efficiently as the product of an upper triangular matrix and a vector. When  $\Sigma$  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  $\Sigma$  or  $\Sigma^{-1}$  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  $\Sigma$  is given, then evaluate  $x = Lz + \mu$ . If  $\Sigma^{-1}$  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:

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

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

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

These algorithms apply when the covariance/precision matrix is either sparse or dense. When the matrix is dense, the computational complexity is  $\mathcal{O}(M^3)$  for a Cholesky decomposition, and  $\mathcal{O}(M^2)$  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  $\Sigma$  (or  $\Sigma^{-1}$ ) 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 class for our purposes is *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 borrows 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  $\beta_i = (\beta_{i1}, \dots, \beta_{ik})$ . 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  $\beta_i$  and  $w_i$ , but does not depend directly on  $\beta_j$  and  $w_j$  for another household  $j \neq i$ .

$$p_i = \frac{\exp(w_i' \beta_i)}{1 + \exp(w_i' \beta_i)}, \quad i = 1 \dots N \quad (6)$$

The coefficient vectors  $\beta_i$  are distributed across the population of households following a MVN 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  $\mu$ , so we place a multivariate normal prior on  $\mu$ , 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  $N$   $\beta_i$  vectors, and the  $k$  elements in  $\mu$ .

The log posterior density, ignoring any normalization constants, is

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

Because one element of  $\beta_i$  can be correlated with another element of  $\beta_i$  (for the same unit), we allow for the cross-partials between elements of  $\beta_i$  for any  $i$  to be nonzero. Also, because the mean of each  $\beta_i$  depends on  $\mu$ , the cross-partials between  $\mu$  and any  $\beta_i$  can be nonzero. However, since the  $\beta_i$  and  $\beta_j$  are independent samples, and the  $y_i$  are conditionally independent, the cross-partial derivatives between an element of  $\beta_i$  and any element of any  $\beta_j$  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  $\beta_i$  for each unit together, and place  $\mu$  at the end.

$$\beta_{11}, \dots, \beta_{1k}, \beta_{21}, \dots, \beta_{2k}, \dots, \beta_{N1}, \dots, \beta_{Nk}, \mu_1, \dots, \mu_k \quad (8)$$

In this case, the Hessian has a “block-arrow” pattern. Figure 1a illustrates this pattern for  $N = 5$  and  $k = 2$  (12 total variables).

Another possibility is to group coefficients for each covariate together.

$$\beta_{11}, \dots, \beta_{N1}, \beta_{12}, \dots, \beta_{N2}, \dots, \beta_{1k}, \dots, \beta_{Nk}, \mu_1, \dots, \mu_k \quad (9)$$

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.<sup>3</sup> If  $N = 1,000$ , then  $M = 2,002$ , with more than 4 million elements

<sup>3</sup>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

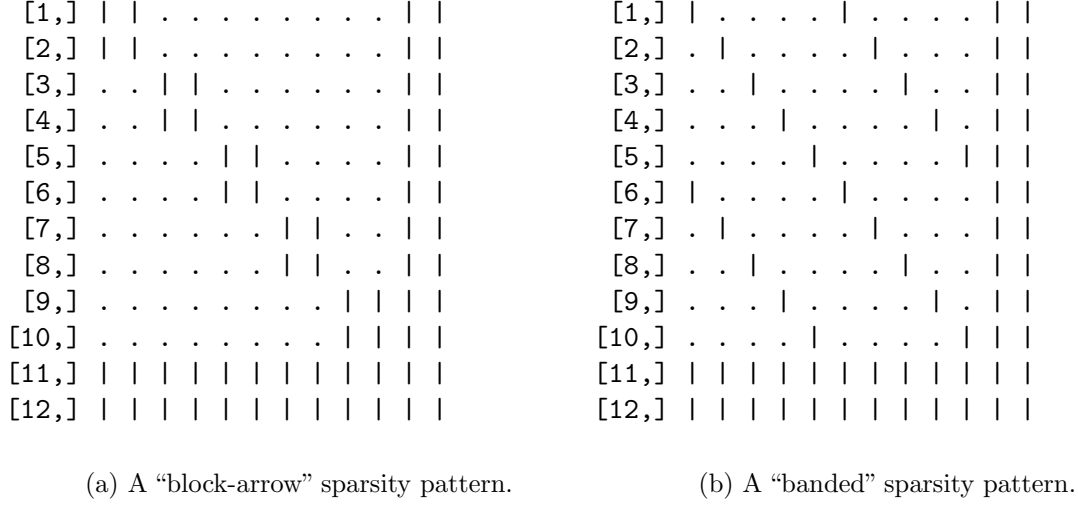


Figure 1: Two examples of sparsity patterns for a hierarchical model.

in the 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 MVN precision matrix  $\Sigma^{-1}$  of a MVN approximation to the posterior distribution of  $(\beta_{1:N}, \mu)$ . If we were to simulate from this MVN using `rmvnorm`, or evaluate MVN densities using `dmvnorm`, we would need to invert  $\Sigma^{-1}$  to  $\Sigma$  first, and store it as a dense matrix. Internally, `rmvnorm` and `dmvnorm` invoke dense linear algebra routines, including matrix factorization.

## 2. Using the sparseMVN package

The signatures of the **sparseMVN** key sparse matrix functions 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`.

Table 1 describes the function arguments. These functions do require the user to compute the Cholesky decomposition `CH` beforehand, but this needs to be done only once (as long as  $\Sigma$  or  $\Sigma^{-1}$  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 to compute `CH`, the **sparseMVN** functions will handle that directly, with no additional user intervention required. The `chol` function in **base R** should not be used.

---

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

---

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

## 2.1. An example

Suppose we want to generate samples from an MVN approximation to the posterior distribution of our example model from Section 1.2. `sparseMVN` 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 <- sparseMVN::binary.sim(N=50, k=2, T=50)
R> priors <- list(inv.A=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=priors,
+                   method="Sparse",
+                   control=list(function.scale.factor=-1))
```

The call to `trust.optim` returns the posterior mode, and the Hessian evaluated at the mode. These results 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)
```

We can then sample from the posterior using an MVN approximation, and compute the MVN log density for each sample.

```
R> samples <- rmvn.sparse(R, pm, CH, prec=TRUE)
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/precision matrices with the same structure as the Hessian of the log posterior density of the example model in Section 2.1. Parameters are ordered such that the matrix has a block-arrow pattern, as in Figure 1a. The size and sparsity of the test matrices vary through manipulation of the number of blocks ( $N$ ), the size of each block ( $k$ ), and the number of rows/columns in the margin (also  $k$ ). Each test matrix has  $(N + 1)k$  rows and columns. Table 2 summarizes the case conditions.

Figure 2 compares mean run times to compute 1,000 MVN densities, and generate 1,000 MVN samples, using functions in **sparseMVN** (**rmvn.sparse**, **dmvn.sparse**) and **mvtnorm** (**rmvnorm**, **dmvnorm**). Times were collected over 300 replications on a 2013-vintage Mac Pro with a 12-core 2.7 GHz Intel Xeon E5 processor and 64 GB of RAM. The times for **mvtnorm** are faster than **sparseMVN** for low dimensional cases ( $N \leq 50$ ), but grow quadratically in the number of blocks.<sup>4</sup> This is because the number of elements stored in a dense covariance matrix grows quadratically with the number of variables, even though the number of *nonzero* elements grows linearly. In this example, storage and computation requirements for the sparse matrix grow linearly with the number of variables, so the **sparseMVN** run times grow linearly as well. The comparative advantage of **sparseMVN** increases with the sparsity of the covariance matrix.<sup>5</sup>

---

<sup>4</sup>As an example, in the  $N = 10$ ,  $k = 2$  case, the mean time to compute 1,000 MVN densities is just over one millisecond using **dmvnorm**, but more than four milliseconds using **dmvn.sparse**.

<sup>5</sup>Across all cases there was hardly any difference in the **sparseMVN** run times when providing the precision matrix instead of the covariance.

				nonzeros		
	$N$	variables	elements	full	lower tri	sparsity
k=2	10	22	484	124	73	0.256
	20	42	1,764	244	143	0.138
	50	102	10,404	604	353	0.058
	100	202	40,804	1,204	703	0.030
	200	402	161,604	2,404	1,403	0.015
	300	602	362,404	3,604	2,103	0.010
	400	802	643,204	4,804	2,803	0.007
	500	1,002	1,004,004	6,004	3,503	0.006
k=4	10	44	1,936	496	270	0.256
	20	84	7,056	976	530	0.138
	50	204	41,616	2,416	1,310	0.058
	100	404	163,216	4,816	2,610	0.030
	200	804	646,416	9,616	5,210	0.015
	300	1,204	1,449,616	14,416	7,810	0.010
	400	1,604	2,572,816	19,216	10,410	0.007
	500	2,004	4,016,016	24,016	13,010	0.006

Table 2: Cases for timing comparison.  $N$  and  $k$  refer, respectively, to the number of blocks in the block-arrow structure (analogous to heterogeneous units in the binary choice example), and the size of each block. The total number of variables is  $M = (N + 1)k$ , and the total number of elements in the matrix is  $M^2$ . The three rightmost columns present the number of nonzeros in the full matrix and lower triangle, and the sparsity (proportion of matrix elements that are nonzero).



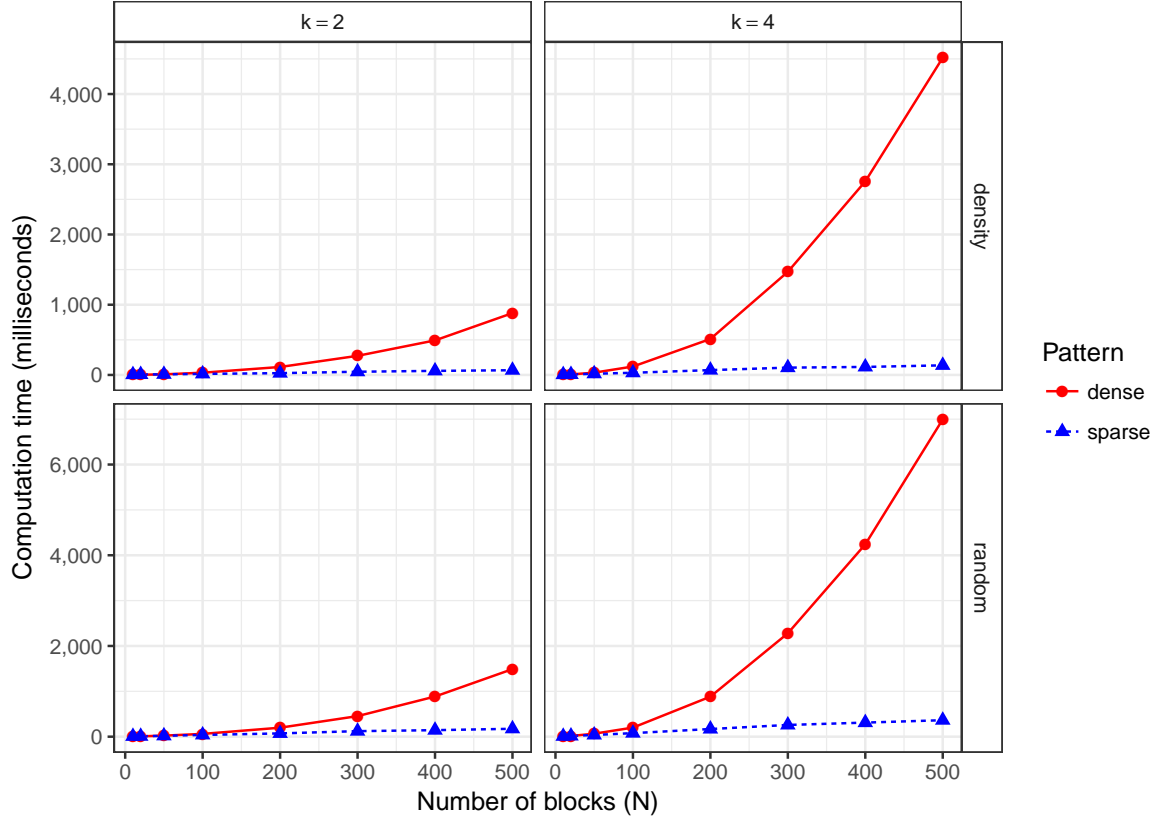


Figure 2: Mean computation time for simulating 1,000 MVN samples, and computing 1,000 MVN densities, averaged over 300 replications. Densities were computed using `dmvnorm` and `dmvn.sparse`, while random samples were generated using `rmvnorm` and `rmvn.sparse`.

The `sparseMVN` functions always require a sparse Cholesky decomposition of the covariance or precision matrix, and the `mvtnorm` functions require a dense precision matrix to be inverted into a dense covariance matrix. Figure 3 compares the computation times of these preparatory steps. There are three cases to consider: inverting a dense matrix using the `solve` function, decomposing a sparse matrix using `Cholesky`, and decomposing a dense matrix using `chol`. Applying `chol` to a dense function is not a required operation in advance of calling `rmvnorm` or `dmvnorm`, but those functions will invoke some kind of decomposition internally. We include it in our comparison because it comprises a substantial part of the computation time. The decomposition and inversion operations on the dense matrices grow cubically as the size of the matrix increases. The sparse Cholesky decomposition time is negligible. For example, the mean run time for the  $N = 500$ ,  $k = 4$  case is about 0.39 milliseconds.

Code to replicate the data used in Figures 2 and 3 is available as an online supplement to this paper, and in the `doc/` directory of the installed package.

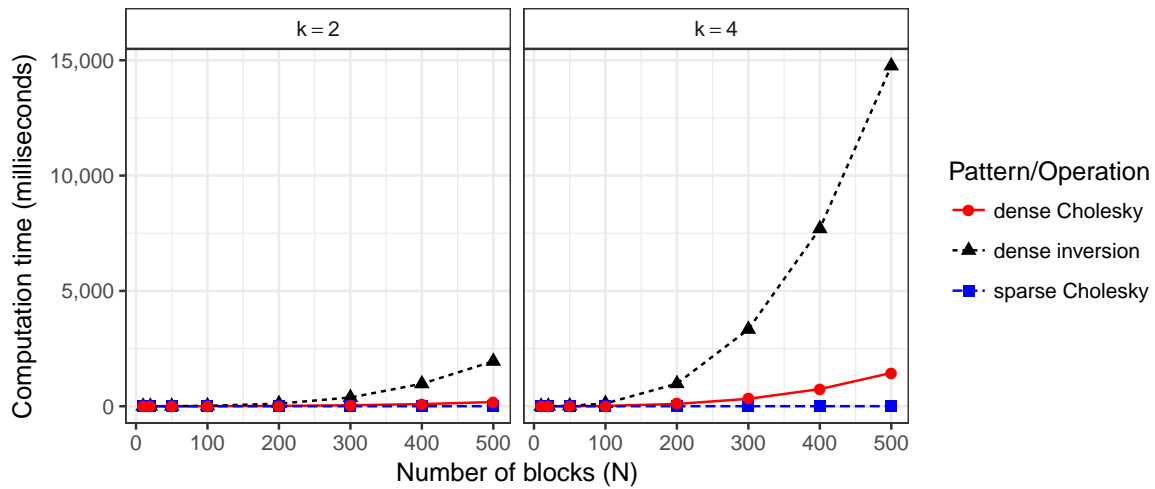


Figure 3: Computation time for Cholesky decomposition of sparse and dense matrices, and inversion of dense matrices.

## 4. Other packages for creating and using sparse matrices

Deriving utility from **sparseMVN** depends on being able to create and use sparse matrices. This section describes briefly two other R packages that may be useful.

### 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. The derivation of the Hessian might be too tedious or complicated to derive it analytically, or too large to estimate numerically. However, for many applications (e.g., a hierarchical model with conditional independence), it is easy to identify which elements in the Hessian are nonzero.

The **sparseHessianFD** package (Braun 2016) provides functions to estimate 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, which may be more stable and robust than line search approaches.

## 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  $\mu$  changes from call to call, but  $\Sigma$  does not.
2. In some applications, the precision matrix  $\Sigma^{-1}$ , and not the covariance matrix  $\Sigma$ , 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,  $\Sigma^{-1}$  would first have to be inverted explicitly.
3. The **mvtnorm** functions treat  $\Sigma$  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  $\Sigma$  or  $\Sigma^{-1}$  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  $\Sigma^{-1}$ , 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  $\Sigma$  over and over. Nevertheless, the main purpose of **sparseMVN** is to exploit sparsity in either  $\Sigma$  or  $\Sigma^{-1}$  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](mailto:braunm@smu.edu)  
 URL: <http://www.smu.edu/Cox/Departments/FacultyDirectory/BraunMichael>