

Timing Rayleigh Quotient minimization in R

true

2023-9-6

Abstract

This vignette is simply to record the methods and results for timing various Rayleigh Quotient minimizations with R using different functions and different ways of running the computations, in particular trying Fortran subroutines and the R byte compiler. It has been updated from a 2012 document to reflect changes in R and its packages that make it awkward to reprocess the original document on newer computers.

The computational task

The maximal and minimal eigensolutions of a symmetric matrix A are extrema of the Rayleigh Quotient

$$R(x) = (x'Ax)/(x'x)$$

We could also deal with generalized eigenproblems of the form

$$Ax = eBx$$

where B is symmetric and positive definite by using the Rayleigh Quotient (RQ)

$$R_g(x) = (x'Ax)/(x'Bx)$$

In this document, B will always be an identity matrix, but some programs we test assume that it is present.

Note that the objective is scaled by the parameters, in fact by their sum of squares. Alternatively, we may think of requiring the **normalized** eigensolution, which is given as

$$x_{normalized} = x/\sqrt{x'x}$$

Timings and speedups

In R, execution times can be measured by the function `system.time`, and in particular the third element of the object this function returns. However, various factors influence computing times in a modern computational system, so we generally want to run replications of the times. The R packages `rbenchmark` and `microbenchmark` can be used for this. I have a preference for the latter. However, to keep the time to prepare this vignette with `Sweave` or `knitr` reasonable, many of the timings will be done with only `system.time`.

There are some ways to speed up R computations.

- The code can be modified to use more efficient language structures. We show some of these below, in particular, to use vector operations.
- We can use the R byte code compiler by Luke Tierney, which has been part of the R distribution since version 2.14.
- We can use compiled code in other languages. Here we show how Fortran subroutines can be used.

Our example matrix

We will use a matrix called the Moler matrix Nash (1979, Appendix 1). This is a positive definite symmetric matrix with one small eigenvalue. We will show a couple of examples of computing the small eigenvalue solution, but will mainly perform timings using the maximal eigenvalue solution, which we will find by minimizing the RQ of (-1) times the matrix. (The eigenvalue of this matrix is the negative of the maximal eigenvalue of the original, but the eigenvectors are equivalent to within a scaling factor for non-degenerate eigenvalues.)

Here is the code for generating the Moler matrix.

```
molermat<-function(n){
  A<-matrix(NA, nrow=n, ncol=n)
  for (i in 1:n){
    for (j in 1:n) {
      if (i == j) A[i,i]<-i
      else A[i,j]<-min(i,j) - 2
    }
  }
  A
}
```

However, since R is more efficient with vectorized code, the following routine by Ravi Varadhan should do much better.

```
molerfast <- function(n) {
  # A fast version of `molermat'
  A <- matrix(0, nrow = n, ncol = n)
  j <- 1:n
  for (i in 1:n) {
    A[i, 1:i] <- pmin(i, 1:i) - 2
  }
  A <- A + t(A)
  diag(A) <- 1:n
  A
}
```

Time to build the matrix

Let us see how long it takes to build the Moler matrix of different sizes. In 2012 we used the byte-code compiler, but that now seems to be active by default and NOT to give worthwhile improvements. We also include times for the `eigen()` function that computes the full set of eigensolutions very quickly.

Loading required package: microbenchmark

##		n	buildi	osize	eigntime	bfast
## 1	50	2933	20216	1272	2683	
## 2	100	6504	80216	1850	672	
## 3	150	16538	180216	3922	976	
## 4	200	27359	320216	6116	1444	
## 5	250	33478	500216	7498	1765	
## 6	300	43275	720216	11065	2568	
## 7	350	53450	980216	13823	2842	
## 8	400	65587	1280216	17847	3789	
## 9	450	81127	1620216	22934	4454	
## 10	500	94564	2000216	29266	4529	

buildi - interpreted build time

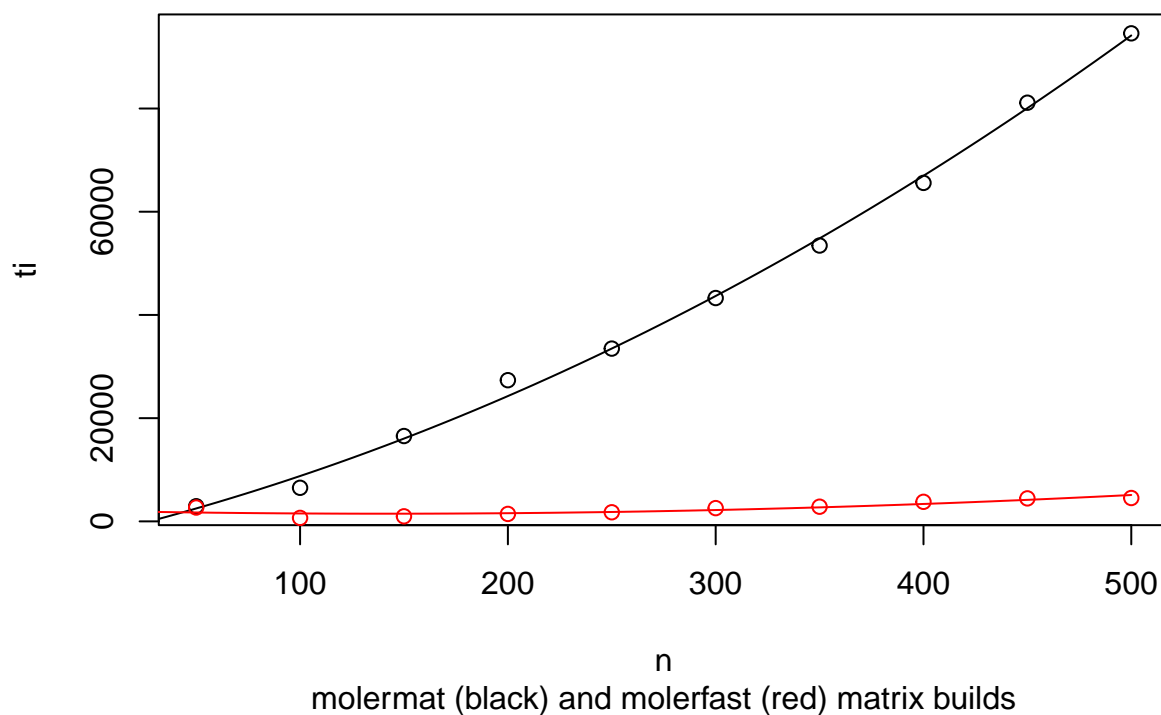
```
## osize - matrix size in bytes; eigentime - all eigensolutions time
## bfast - interpreted vectorized build time
## Times converted to milliseconds
```

It does not appear that the compiler has much effect, or else it is being automatically invoked.

We can graph the times. The code, which is not echoed here, also models the times and the object size created as almost perfect quadratic models in n .

```
##
## Call:
## lm(formula = ti ~ n + n2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2283.0 -1134.4   196.5   476.1  3077.7
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.807e+03  2.006e+03  -1.399  0.204445
## n              9.644e+01  1.675e+01   5.756  0.000694 ***
## n2             1.950e-01  2.969e-02   6.568  0.000314 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1705 on 7 degrees of freedom
## Multiple R-squared:  0.9977, Adjusted R-squared:  0.997
## F-statistic: 1493 on 2 and 7 DF, p-value: 6.193e-10
##
## Call:
## lm(formula = tf ~ n + n2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
##  -863.21  -420.87    35.16   348.58   961.90
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2048.26982  730.85442   2.803   0.0264 *
## n           -7.94512    6.10471  -1.301   0.2343
## n2            0.02814    0.01082   2.602   0.0353 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 621.4 on 7 degrees of freedom
## Multiple R-squared:  0.8412, Adjusted R-squared:  0.7959
## F-statistic: 18.54 on 2 and 7 DF, p-value: 0.001595
```

Execution time vs matrix size



```
## Warning in summary.lm(osize): essentially perfect fit: summary may be
## unreliable
```

```
##
```

```
## Call:
```

```
## lm(formula = os ~ n + n2)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -2.654e-12 -1.314e-13  3.293e-13  7.262e-13  1.211e-12
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)  2.160e+02  1.617e-12  1.336e+14 < 2e-16 ***
## n            5.127e-13  1.351e-14  3.795e+01 2.29e-09 ***
## n2           8.000e+00  2.394e-17  3.342e+17 < 2e-16 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

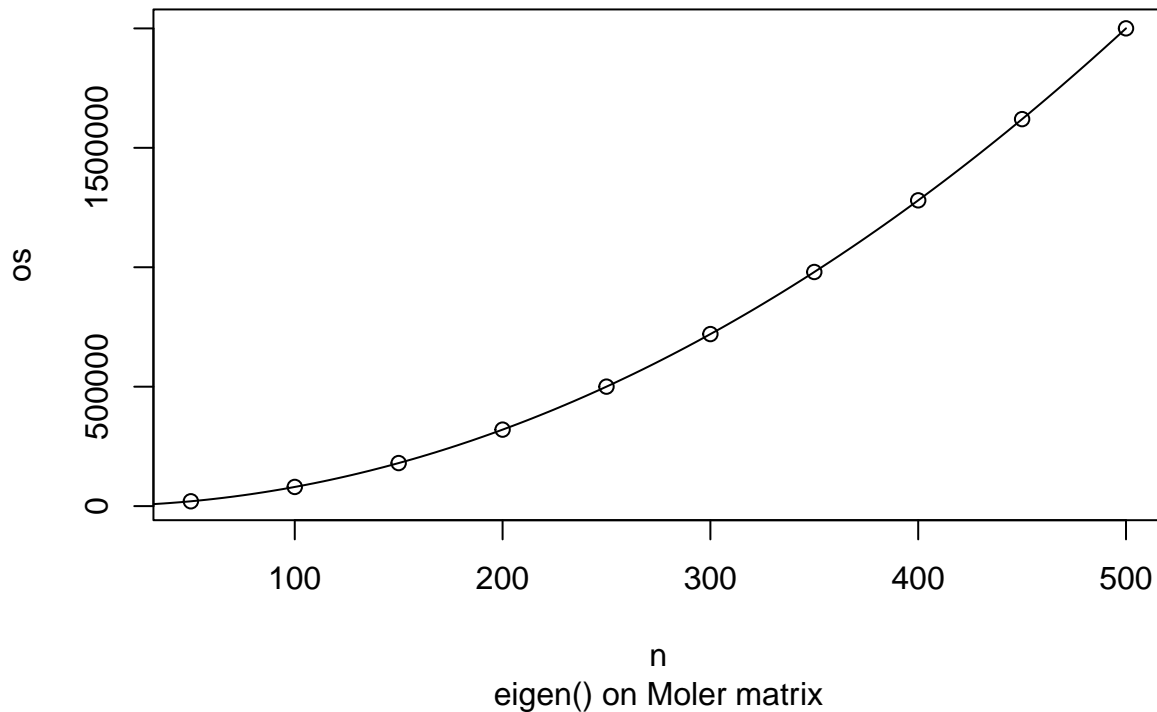
```
##
```

```
## Residual standard error: 1.375e-12 on 7 degrees of freedom
```

```
## Multiple R-squared:  1, Adjusted R-squared:  1
```

```
## F-statistic: 1.112e+36 on 2 and 7 DF, p-value: < 2.2e-16
```

Execution time vs matrix size



Computing the Rayleigh Quotient

The Rayleigh Quotient requires the quadratic form $x'Ax$ divided by the inner product $x'x$. R lets us form this in several ways.

```
rqdir<-function(x, AA){  
  rq<-0.0  
  n<-length(x) # assume x, AA conformable  
  for (i in 1:n) {  
    for (j in 1:n) {  
      rq<-rq+x[i]*AA[[i,j]]*x[j]  
    }  
  }  
  rq  
}
```

Somewhat better (as we shall show below) is

```
ray1<-function(x, AA){  
  rq<- t(x)%*%AA%*%x  
}
```

and (believed) better still is

```
ray2<-function(x, AA){  
  rq<- as.numeric(crossprod(x, crossprod(AA,x)))  
}
```

Note that we could implicitly include the minus sign in these routines to allow for finding the maximal eigenvalue by minimizing the Rayleigh Quotient of $-A$. However, such shortcuts often rebound when the

implicit negation is overlooked.

If we already have the inner product $x^T A x$ as vector `ax` from some other computation, then we can simply use

```
ray3<-function(x, AA, ax=axftn){
  # ax is a function to form AA%%x
  rq<- - as.numeric(crossprod(x, ax(x, AA)))
}
```

Matrix-vector products

In generating the RQ, we do not actually need the matrix itself, but simply the inner product with a vector `x`, from which a second inner product with `x` gives us the quadratic form $x^T A x$. If `n` is the order of the problem, then for large `n`, we avoid storing and manipulating a very large matrix if we use **implicit inner product** formation. We do this with the following code. For future reference, we include the multiplication by an identity.

```
ax<-function(x, AA){
  u<- as.numeric(AA%%x)
}

axx<-function(x, AA){
  u<- as.numeric(crossprod(AA, x))
}
```

Note that second argument, supposedly communicating the matrix which is to be used in the matrix-vector product, is ignored in the following implicit product routine. It is present only to provide a common syntax when we wish to try different routines within other computations.

```
aximp<-function(x, AA=1){ # implicit moler A*x
  n<-length(x)
  y<-rep(0,n)
  for (i in 1:n){
    tt<-0.
    for (j in 1:n) {
      if (i == j) tt<-tt+i*x[i]
      else tt<-tt+(min(i,j) - 2)*x[j]
    }
    y[i]<-tt
  }
  y
}

ident<-function(x, B=1) x # identity
```

However, Ravi Varadhan has suggested the following vectorized code for the implicit matrix-vector product.

```
axmolerfast <- function(x, AA=1) {
  # A fast and memory-saving version of A%%x
  # For Moler matrix. Note we need a matrix argument to match other functions
  n <- length(x)
  j <- 1:n
  ax <- rep(0, n)
  for (i in 1:n) {
    term <- x * (pmin(i, j) - 2)
    ax[i] <- sum(term[-i])
  }
}
```

```
ax <- ax + j*x
ax
}
```

We can also use external language routines, for example in Fortran. However, this needs a Fortran **subroutine** which outputs the result as one of the returned components. The subroutine is in file `moler.f`.

```
subroutine moler(n, x, ax)
integer n, i, j
double precision x(n), ax(n), sum
c return ax = A * x for A = moler matrix
c A[i,j]=min(i,j)-2 for i<>j, or i for i==j
do 20 i=1,n
sum=0.0
do 10 j=1,n
if (i.eq.j) then
sum = sum+i*x(i)
else
sum = sum+(min(i,j)-2)*x(j)
endif
10 continue
ax(i)=sum
20 continue
return
end
```

This is then compiled in a form suitable for R use by the command (this is a command-line tool, and was run in Ubuntu Linux in a directory containing the file `moler.f` but outside this vignette):

```
R CMD SHLIB moler.f
```

This creates files `moler.o` and `moler.so`, the latter being the dynamically loadable library we need to bring into our R session.

```
dyn.load("moler.so")
cat("Is the mat multiply loaded? ",is.loaded("moler"),"\n")

## Is the mat multiply loaded? TRUE

axftn<-function(x, AA=1) { # ignore second argument
n<-length(x) # could speed up by having this passed
vout<-rep(0,n) # purely for storage
res<-(.Fortran("moler", n=as.integer(n), x=as.double(x), vout=as.double(vout)))$vout
}
```

We can also byte compile each of the routines above

Now it is possible to time the different approaches to the matrix-vector product.

```
dyn.load("moler.so")
cat("Is the mat multiply loaded? ",is.loaded("moler"),"\n")

## Is the mat multiply loaded? TRUE

require(microbenchmark)
nmax<-10
ptable<-matrix(NA, nrow=nmax, ncol=6) # to hold results
# loop over sizes
for (ni in 1:nmax){
```

```

n<-50*ni
x<-runif(n) # generate a vector
ptable[[ni, 1]]<-n
AA<-molermat(n)
tax<- mean(microbenchmark(oax<-ax(x, AA), times=mbt)$time)*0.001
taxx<-mean(microbenchmark(oaxx<-axx(x, AA), times=mbt)$time)*0.001
if (! identical(oax, oaxx)) stop("oaxx NOT correct")
taxftn<-mean(microbenchmark(oaxftn<-axftn(x, AA=1), times=mbt)$time)*0.001
if (! identical(oax, oaxftn)) stop("oaxftn NOT correct")
taximp<-mean(microbenchmark(oaximp<-aximp(x, AA=1), times=mbt)$time)*0.001
if (! identical(oax, oaximp)) stop("oaximp NOT correct")
taxmfi<-mean(microbenchmark(oaxmfi<-axmolerfast(x, AA=1), times=mbt)$time)*0.001
if (! identical(oax, oaxmfi)) stop("oaxmfi NOT correct")
ptable[[ni, 2]]<-tax
ptable[[ni, 3]]<-taxx
ptable[[ni, 4]]<-taxftn
ptable[[ni, 5]]<-taximp
ptable[[ni, 6]]<-taxmfi
}

axtym<-data.frame(n=ptable[,1], ax=ptable[,2], axx=ptable[,3], axftn=ptable[,4],
                  aximp=ptable[,5], axmfast=ptable[,6])
print(axtym)

```

```

##      n      ax      axx      axftn      aximp axmfast
## 1   50 627.8485 637.1915 1163.7515 3712.568 2784.230
## 2  100  79.6785 4997.6255  46.7465 4303.340  644.580
## 3  150  86.3265  38.3855  86.6805 8426.368 1180.360
## 4  200  82.5710  63.8790 141.9605 13880.938 1352.947
## 5  250 104.1345  82.6410 212.8070 20821.134 1788.096
## 6  300 123.3020 107.8635 299.1515 29349.627 2233.501
## 7  350 168.2050 100.8210 399.4620 39050.430 2746.294
## 8  400 177.1740 127.1790 514.4995 50699.613 3232.972
## 9  450 200.6760 163.0270 647.5695 62902.316 3831.508
## 10 500 227.7360 202.2215 825.4050 77665.072 4446.244

```

```

# explain table

expln <- c("ax = R matrix * vector  A %*% x",
          "axx = R crossprod A, x",
          "axftn = Fortran version of implicit Moler A * x",
          "aximp = implicit moler A*x in R",
          "axmfast = A fast and memory-saving version of A %*% x")
for (exx in expln) { cat(exx, "\n")}

```

```

## ax = R matrix * vector  A %*% x
## axx = R crossprod A, x
## axftn = Fortran version of implicit Moler A * x
## aximp = implicit moler A*x in R
## axmfast = A fast and memory-saving version of A %*% x

```

From the above output, we see that the `crossprod` variant of the matrix-vector product appears to be the fastest. However, we have omitted the time to build the matrix. If we must build the matrix, then we need somehow to include that time. Because the times for the matrix-vector product were so short, we used `replicate` above to run 100 copies of the same calculation, which may give some distortion of the timings.

However, we believe the scale of the times is more or less correct. To compare these times to the times for the Fortran or implicit matrix-vector routines, we should add a multiple of the relevant interpreted or compiled build times. Here we have used the times for the rather poor `moler`mat() function, but this is simply to illustrate the range of potential timings. Apportioning such “fixed costs” to timings is never a trivial decision. Similarly if, where and how to store large matrices if we do build them, and whether it is worth building them more than once if storage is an issue, are all questions that may need to be addressed if performance becomes important.

```
adjtym<-data.frame(n=axtym$n, axx1=axtym$axx+1*bmattym$buildi,
  axxz=axtym$axx+100*bmattym$buildi,
  axftn=axtym$axftn, aximp=axtym$aximp)
print(adjtym)
```

##	n	axx1	axxz	axftn	aximp
## 1	50	3570.224	293940.5	1163.7515	3712.568
## 2	100	11502.084	655443.4	46.7465	4303.340
## 3	150	16576.618	1653861.6	86.6805	8426.368
## 4	200	27422.530	2735928.9	141.9605	13880.938
## 5	250	33560.363	3347854.8	212.8070	20821.134
## 6	300	43382.927	4327614.2	299.1515	29349.627
## 7	350	53550.490	5345067.7	399.4620	39050.430
## 8	400	65714.074	6558816.7	514.4995	50699.613
## 9	450	81290.419	8112902.2	647.5695	62902.316
## 10	500	94765.848	9456564.8	825.4050	77665.072

Out of all this, we see that the Fortran implicit matrix-vector product is the overall winner at all values of `n`. Moreover, it does NOT require the creation and storage of the matrix. However, using Fortran does involve rather more work for the user, and for most applications it is likely we could live with the use of either

- the interpreted matrix-product based on `crossprod` and an actual matrix is good enough, especially if a fast matrix build is used and we have plenty of memory, or
- the interpreted or byte-code compiled implicit matrix-vector multiply `axmolerfast`.

RQ computation times

We have set up three versions of a Rayleigh Quotient calculation in addition to the direct form. The third form is set up to use the `axftn` routine that we have already shown is efficient. We could also use this with the implicit matrix-vector product `axmolerfast`.

It seems overkill to show the RQ computation time for all versions and matrices, so we will do the timing simply for a matrix of order 500.

```
dyn.load("moler.so")
n<-500
x<-runif(n) # generate a vector
AA<-moler
```

mat(n)
tdi<-microbenchmark(rdi<-rqdir(x, AA))\$time
cat("Direct algorithm: ",mean(tdi)*0.001,"\n")

```
## Direct algorithm: 15659.69
```

```
t1i<-microbenchmark(r1i<-ray1(x, AA))$time
cat("ray1: mat-mult algorithm: ", mean(t1i)*0.001,"\n")
```

```
## ray1: mat-mult algorithm: 165.6404
```

```
t2i<-microbenchmark(r2i<-ray2(x, AA))$time
cat("ray2: crossprod algorithm: ",mean(t2i)*0.001,"\n")
```

```
## ray2: crossprod algorithm: 175.2721
```

```
t3fi<-microbenchmark(r3i<-ray3(x, AA, ax=axftn))$time  
cat("ray3: ax Fortran + crossprod: ",mean(t3fi)*0.001,"\n")
```

```
## ray3: ax Fortran + crossprod: 719.3521
```

```
t3ri<-microbenchmark(r3i<-ray3(x, AA, ax=axmolerfast))$time  
cat("ray3: ax fast R implicit + crossprod: ",mean(t3ri)*0.001,"\n")
```

```
## ray3: ax fast R implicit + crossprod: 4846.453
```

Here we see that the use of the `crossprod` in `ray2` is very fast, and this is interpreted code. Once again, we note that all timings except those for `ray3` should have some adjustment for the building of the matrix. If storage is an issue, then `ray3`, which uses the implicit matrix-vector product in Fortran, is the approach of choice. My own preference would be to use this option if the Fortran matrix-vector product subroutine is already available for the matrix required. I would not, however, generally choose to write the Fortran subroutine for a “new” problem matrix. The fast implicit matrix-vector tool with `ray3` is also useful and quite fast if we need to minimize memory use.

Solution by spg

To actually solve the eigensolution problem we will first use the projected gradient method `spg` from `BB`. We repeat the `RQ` function so that it is clear which routine we are using.

```
rqt<-function(x, AA){  
  rq<-as.numeric(crossprod(x, crossprod(AA,x)))  
}  
proj<-function(x) { sign(x[1])*x/sqrt(crossprod(x)) }  
require(BB)
```

```
## Loading required package: BB
```

```
n<-100  
x<-rep(1,n)  
AA<-molermat(n)  
evs<-eigen(AA)  
tmin<-microbenchmark(amin<-spg(x, fn=rqt, project=proj, control=list(trace=FALSE), AA=AA), times=mbt)$t
```

```
## Warning in sign(x[1]) * x/sqrt(crossprod(x)): Recycling array of length 1 in vector-array arithmetic  
## Use c() or as.vector() instead.
```

```
## Warning in sign(x[1]) * x/sqrt(crossprod(x)): Recycling array of length 1 in vector-array arithmetic  
## Use c() or as.vector() instead.
```

```
## Warning in spg(x, fn = rqt, project = proj, control = list(trace = FALSE), :  
## convergence tolerance satisfied at intial parameter values.
```

```
## Warning in sign(x[1]) * x/sqrt(crossprod(x)): Recycling array of length 1 in vector-array arithmetic  
## Use c() or as.vector() instead.
```

```
## Warning in sign(x[1]) * x/sqrt(crossprod(x)): Recycling array of length 1 in vector-array arithmetic  
## Use c() or as.vector() instead.
```

```
## Warning in spg(x, fn = rqt, project = proj, control = list(trace = FALSE), :  
## convergence tolerance satisfied at intial parameter values.
```

```
#amin  
tmax<-microbenchmark(amax<-spg(x, fn=rqt, project=proj, control=list(trace=FALSE), AA=-AA), times=mbt)$
```

```

## Warning in sign(x[1]) * x/sqrt(crossprod(x)): Recycling array of length 1 in vector-array arithmetic
## Use c() or as.vector() instead.

## Warning in sign(x[1]) * x/sqrt(crossprod(x)): Recycling array of length 1 in vector-array arithmetic
## Use c() or as.vector() instead.

## Warning in spg(x, fn = rqt, project = proj, control = list(trace = FALSE), :
## convergence tolerance satisfied at initial parameter values.

## Warning in sign(x[1]) * x/sqrt(crossprod(x)): Recycling array of length 1 in vector-array arithmetic
## Use c() or as.vector() instead.

## Warning in sign(x[1]) * x/sqrt(crossprod(x)): Recycling array of length 1 in vector-array arithmetic
## Use c() or as.vector() instead.

## Warning in spg(x, fn = rqt, project = proj, control = list(trace = FALSE), :
## convergence tolerance satisfied at initial parameter values.

#amax
evalmax<-evs$values[1]
evecmax<-evs$vectors[,1]
evecmax<-sign(evecmax[1])*evecmax/sqrt(as.numeric(crossprod(evecmax)))
emax<-list(evalmax=evalmax, evecmax=evecmax)
# save(emax, file="temax.Rdata")
evalmin<-evs$values[n]
evecmin<-evs$vectors[,n]
evecmin<-sign(evecmin[1])*evecmin/sqrt(as.numeric(crossprod(evecmin)))
avecmax<-amax$par
avecmin<-amin$par
avecmax<-sign(avecmax[1])*avecmax/sqrt(as.numeric(crossprod(avecmax)))
avecmin<-sign(avecmin[1])*avecmin/sqrt(as.numeric(crossprod(avecmin)))
cat("minimal eigensolution: Value=",amin$value,"in time ",tmin,"\n")

## minimal eigensolution: Value= 318550 in time 9220733 1734173
cat("Eigenvalue - result from eigen=",amin$value-evalmin," vector max(abs(diff))=",
    max(abs(avecmin-evecmin)),"\n\n")

## Eigenvalue - result from eigen= 318550 vector max(abs(diff))= 0.7660254
#print(amin$par)
cat("maximal eigensolution: Value=",amax$value,"in time ",tmax,"\n")

## maximal eigensolution: Value= 318550 in time 1947695 1734030
cat("Eigenvalue - result from eigen=",amax$value-evalmax," vector max(abs(diff))=",
    max(abs(avecmax-evecmax)),"\n\n")

## Eigenvalue - result from eigen= 314615.7 vector max(abs(diff))= 0.242489
#print(amax$par)

## Loading required package: compiler

## Warning in sign(x[1]) * x/sqrt(crossprod(x)): Recycling array of length 1 in vector-array arithmetic
## Use c() or as.vector() instead.

## Warning in sign(x[1]) * x/sqrt(crossprod(x)): Recycling array of length 1 in vector-array arithmetic
## Use c() or as.vector() instead.

```



```
## Warning in sign(x[1]) * x/sqrt(crossprod(x)): Recycling array of length 1 in vector-array arithmetic
## Use c() or as.vector() instead.

## Warning in sign(x[1]) * x/sqrt(crossprod(x)): Recycling array of length 1 in vector-array arithmetic
## Use c() or as.vector() instead.

## Warning in sign(x[1]) * x/sqrt(crossprod(x)): Recycling array of length 1 in vector-array arithmetic
## Use c() or as.vector() instead.

## Warning in sign(x[1]) * x/sqrt(crossprod(x)): Recycling array of length 1 in vector-array arithmetic
## Use c() or as.vector() instead.

## Warning in sign(x[1]) * x/sqrt(crossprod(x)): Recycling array of length 1 in vector-array arithmetic
## Use c() or as.vector() instead.

## Warning in sign(x[1]) * x/sqrt(crossprod(x)): Recycling array of length 1 in vector-array arithmetic
## Use c() or as.vector() instead.

##      n spgrqt tbld
## 1   50   4181  313
## 2  100  19137  653
## 3  150  38945 1844
## 4  200  73055 2670
## 5  250 123846 2594
## 6  300 201101 2229
## 7  350 299590 2792
## 8  400 432109 3207
## 9  450 615891 4119
## 10 500 802472 4465
```

Solution by other optimizers}

We can try other optimizers, but we must note that unlike `spg` they do not take account of the scaling. However, we can build in a transformation, since our function is always the same for all sets of parameters scaled by the square root of the parameter inner product. The function `nobj` forms the quadratic form that is the numerator of the Rayleigh Quotient using the more efficient `code{crossprod()}` function

```
rq<- as.numeric(crossprod(y, crossprod(AA,y)))
```

but we first form

```
y<-x/sqrt(as.numeric(crossprod(x)))
```

to scale the parameters.

Since we are running a number of gradient-based optimizers in the wrapper `optimx::opm()`, we have reduced the matrix sizes and numbers.

```
require(optimx)
```

```
## Loading required package: optimx
```

```
nobj<-function(x, AA=-AA){
  y<-x/sqrt(as.numeric(crossprod(x)))
  rq<- as.numeric(crossprod(y, crossprod(AA,y)))
}
```

```
ngrobj<-function(x, AA=-AA){
```

```

y<-x/sqrt(as.numeric(crossprod(x)))
n<-length(x)
dd<-sqrt(as.numeric(crossprod(x)))
T1<-diag(rep(1,n))/dd
T2<- x%o%x/(dd*dd*dd)
gt<-T1-T2
gy<- as.vector(2.*crossprod(AA,y))
gg<-as.numeric(crossprod(gy, gt))
}
# require(optplus)
# mset<-c("L-BFGS-B", "BFGS", "CG", "spg", "ucminf", "nlm", "nlminb", "Rummin", "Rcgmin")
mset<-c("L-BFGS-B", "BFGS", "nbg", "spg", "ucminf", "nlm", "nlminb", "nvm")
nmax<-5
for (ni in 1:nmax){
  n<-20*ni
  x<-runif(n) # generate a vector
  AA<-molerfast(n) # make sure defined
  aall<-opm(x, fn=nobj, gr=ngrobj, method=mset, AA=-AA,
    control=list(starttests=FALSE, dowarn=FALSE))
  # optansout(aall, NULL)
  summary(aall, order=value, )
  cat("Above for n=",n," \n")
}

```

```

## Above for n= 20
## Above for n= 40
## Above for n= 60
## Above for n= 80
## Above for n= 100

```

The timings for these matrices of order 20 to 100 are likely too short to be very reliable in detail, but do show that the RQ problem using the scaling transformation and with an analytic gradient can be solved very quickly, especially by the limited memory methods such as L-BFGS-B and nbg. Below we use the latter to show the times over different matrix sizes.

```

ctable<-matrix(NA, nrow=10, ncol=2)
nmax<-10
for (ni in 1:nmax){
  n<-50*ni
  x<-runif(n) # generate a vector
  AA<-molerfast(n) # define matrix
  tcgu<-microbenchmark(arcgu<-optimr(x, fn=nobj, gr=ngrobj, method="nbg",
    AA=-AA), times=mbt)
  ctable[[ni,1]]<-n
  ctable[[ni,2]]<-mean(tcgu$time)*0.001
}
cgtime<-data.frame(n=ctable[,1], tcgmin=ctable[,2])
print(round(cgtime,0))

```

```

##      n tcgmin
## 1   50    934
## 2  100   1235
## 3  150   2009
## 4  200   3880
## 5  250   3902

```

```
## 6 300 17364
## 7 350 12234
## 8 400 16222
## 9 450 34901
## 10 500 16451
```

A specialized minimizer - Geradin's method}

For comparison, let us try the Geradin routine (Appendix 1) as implemented in R by one of us (JN).

```
cat("Test geradin with explicit matrix multiplication\n")
```

```
## Test geradin with explicit matrix multiplication
```

```
n<-10
AA<-molermat(n)
BB=diag(rep(1,n))
x<-runif(n)
tg<-microbenchmark(ag<-geradin(x, ax, bx, AA=AA, BB=BB,
  control=list(trace=FALSE)), times=mbt)
cat("Minimal eigensolution\n")
```

```
## Minimal eigensolution
```

```
print(ag)
```

```
## $x
## [1] -140240.3912 -70120.4965 -35061.0005 -17532.0801 -8769.2373
## [6] -4391.0319 -2208.3520 -1129.8527 -616.2823 -410.8543
##
## $RQ
## [1] 8.582807e-06
##
## $ipr
## [1] 64
##
## $msg
## [1] "Rayleigh Quotient increased in step"
```

```
cat("Geradin time=",mean(tg$time),"n")
```

```
## Geradin time= 30668182
```

```
tgn<-microbenchmark(agn<-geradin(x, ax, bx, AA=-AA, BB=BB,
  control=list(trace=FALSE)), times=mbt)
cat("Maximal eigensolution (negative matrix)\n")
```

```
## Maximal eigensolution (negative matrix)
```

```
print(agn)
```

```
## $x
## [1] -7470488.9 252468.3 7966893.2 15412073.6 22336396.5 28505851.5
## [7] 33711939.4 37778718.2 40568749.4 41987742.7
##
## $RQ
## [1] -31.58981
##
## $ipr
```

```

## [1] 40
##
## $msg
## [1] "Small gradient -- done"

cat("Geradin time=",mean(tgn$time),"\n")

## Geradin time= 535965.5

Let us time this routine with different matrix vector approaches.

naximp<-function(x, A=1){ # implicit moler A*x
  n<-length(x)
  y<-rep(0,n)
  for (i in 1:n){
    tt<-0.
    for (j in 1:n) {
      if (i == j) tt<-tt+i*x[i]
      else tt<-tt+(min(i,j) - 2)*x[j]
    }
    y[i]<- -tt # include negative sign
  }
  y
}

dyn.load("moler.so")
cat("Is the mat multiply loaded? ",is.loaded("moler"),"\n")

## Is the mat multiply loaded? TRUE

naxftn<-function(x, A) { # ignore second argument
  n<-length(x) # could speed up by having this passed
  vout<-rep(0,n) # purely for storage
  res<-(-1)*(.Fortran("moler", n=as.integer(n), x=as.double(x), vout=as.double(vout)))$vout
}

require(microbenchmark)
nmax<-10
gtable<-matrix(NA, nrow=nmax, ncol=6) # to hold results
# loop over sizes
for (ni in 1:nmax){
  n<-50*ni
  x<-runif(n) # generate a vector
  gtable[[ni, 1]]<-n
  AA<-molermat(n)
  BB<-diag(rep(1,n))
  tgax<-microbenchmark(ogax<-geradin(x, ax, bx, AA=-AA, BB=BB, control=list(trace=FALSE)), times=mbt)
  gtable[[ni, 2]]<-mean(tgax$time)
  tgaximp<-microbenchmark(ogaximp<-geradin(x, naximp, ident, AA=1, BB=1, control=list(trace=FALSE)), times=mbt)
  gtable[[ni, 3]]<-mean(tgaximp$time)
  tgaxftn<-microbenchmark(ogaxftn<-geradin(x, naxftn, ident, AA=1, BB=1, control=list(trace=FALSE)), times=mbt)
  gtable[[ni, 4]]<-mean(tgaxftn$time)
}

gtym<-data.frame(n=gtable[,1], ax=gtable[,2], aximp=gtable[,3], axftn=gtable[,4])

```

```
print(gtym)
```

```
##      n      ax      aximp  axftn
## 1   50  803278.5  31047352 2364328
## 2  100 2091183.0 129963260 1637017
## 3  150 2307232.0 252014967 2362744
## 4  200 3467184.0 500349046 4589828
## 5  250 4485761.0 765291691 6077945
## 6  300 8227366.0 1127967859 8383803
## 7  350 7395218.5 1296537064 9375583
## 8  400 9130419.5 1800166450 13468102
## 9  450 10249001.5 2014867720 14631162
## 10 500 13130819.0 2715422574 19891098
```

Let us check that the solution for $n = 100$ by Geradin is consistent with the answer via `eigen()`.

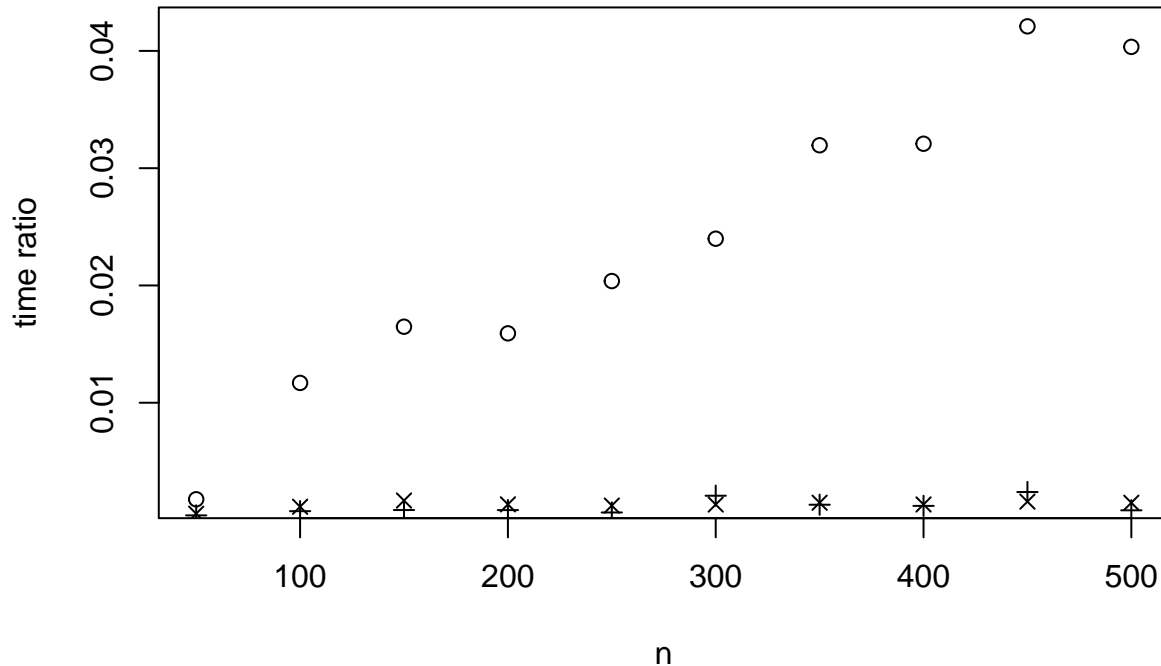
```
n<-100
x<-runif(n)
# emax<-load("temax.Rdata")
evalmax<-emax$evalmax
evecmac<-emax$evecmac
ogaxftn<-geradin(x, naxftn, ident, AA=1, BB=1, control=list(trace=FALSE))
gvec<-ogaxftn$x
gval<-ogaxftn$RQ
gvec<-sign(gvec[[1]])*gvec/sqrt(as.numeric(crossprod(gvec)))
diff<-gvec-evecmax
cat("Geradin diff eigenval from eigen result: ",gval-evalmax,"    max(abs(vector diff))=",
    max(abs(diff)), "\n")
```

```
## Geradin diff eigenval from eigen result:  -1.725128e-06    max(abs(vector diff))= 5.463347e-06
```

Perspective}

We can compare the different approaches by looking at the ratio of the best solution time for each method (compiled or interpreted, with best choice of function) to the time for the Geradin approach for the different matrix sizes. In this we will ignore the fact that some approaches do not build the matrix.

Ratio of eigensolution times to Geradin routine by matrix size



To check the value of the Geradin approach, let us use a much larger problem, with $n=2000$.

```
## Times in seconds
```

```
## Build = 107545393  eigen(): 1135455628  Rcgminu: 394538538  Geradin: 331269167
```

```
## Ratios: build= 0.3246466  eigen= 3.427592  Rcgminu= 1.190991
```

Conclusions}

The Rayleigh Quotient minimization approach to eigensolutions has an intuitive appeal and seemingly offers an interesting optimization test problem, especially if we can make it computationally efficient. To improve time efficiency, we can apply the R byte code compiler, use a Fortran (or other compiled language) subroutine, and choose how we set up our objective functions and gradients. To improve memory use, we can consider using a matrix implicitly.

From the tests in this vignette, here is what we may say about these attempts, which we caution are based on a relatively small sample of tests:

- The R byte code compiler offers a useful gain in speed when our code has statements that access array elements rather than uses them in vectorized form.}
- The `crossprod()` function is very efficient.
- Fortran is not very difficult to use for small subroutines that compute a function such as the implicit matrix-vector product, and it allows efficient computations for such operations.
- The `eigen()` routine is a highly effective tool for computing all eigensolutions, even of a large matrix. It is only worth computing a single solution when the matrix is very large, in which case a specialized method such as that of Geradin makes sense and offers significant savings, especially when combined with the Fortran implicit matrix-product routine.

Acknowledgements

This vignette originated due to a problem suggested by Gabor Grothendieck. Ravi Varadhan has provided inciteful comments and some vectorized functions which greatly altered some of the observations.

Appendix 1: Geradin routine

```
ax<-function(x, AA){
  u<-as.numeric(AA%*%x)
}
bx<-function(x, BB){
  v<-as.numeric(BB%*%x)
}
geradin<-function(x, ax, bx, AA, BB, control=list(trace=TRUE, maxit=1000)){
  # Geradin minimize Rayleigh Quotient, Nash CMN Alg 25
  # print(control)
  trace<-control$trace
  n<-length(x)
  tol<-n*n*.Machine$double.eps^2
  offset<-1e+5 # equality check offset
  if (trace) cat("geradin.R, using tol=",tol,"\n")
  ipr<-0 # counter for matrix mults
  pa<- .Machine$double.xmax
  R<-pa
  msg<-"no msg"
  # step 1 -- main loop
  keepgoing<-TRUE
  while (keepgoing) {
    avec<-ax(x, AA); bvec<-bx(x, BB); ipr<-ipr+1
    xax<-as.numeric(crossprod(x, avec));
    xbx<-as.numeric(crossprod(x, bvec));
    if (xbx <= tol) {
      keepgoing<-FALSE # not really needed
      msg<-"avoid division by 0 as xbx too small"
      break
    }
    p0<-xax/xbx
    if (p0>pa) {
      keepgoing<-FALSE # not really needed
      msg<-"Rayleigh Quotient increased in step"
      break
    }
    pa<-p0
    g<-2*(avec-p0*bvec)/xbx
    gg<-as.numeric(crossprod(g)) # step 6
    if (trace) cat("Before loop: RQ=",p0," after ",ipr," products, gg=",gg,"\n")
    if (gg<tol) { # step 7
      keepgoing<-FALSE # not really needed
      msg<-"Small gradient -- done"
      break
    }
  }
  t<- -g # step 8
  for (itn in 1:n) { # major loop step 9
    y<-ax(t, AA); z<-bx(t, BB); ipr<-ipr+1 # step 10
    tat<-as.numeric(crossprod(t, y)) # step 11
    xat<-as.numeric(crossprod(x, y))
    xbt<-as.numeric(crossprod(x, z))
    tbt<-as.numeric(crossprod(t, z))
    u<-tat*xbt-xat*tbt
```

```

v<-tat*xbx-xax*tbt
w<-xat*xbx-xax*xbt
d<-v*v-4*u*w
if (d<0) stop("Geradin: imaginary roots not possible") # step 13
d<-sqrt(d) # step 14
if (v>0) k<--2*w/(v+d) else k<-0.5*(d-v)/u
xlast<-x # NOT as in CNM -- can be avoided with loop
avec<-avec+k*y; bvec<-bvec+k*z # step 15, update
x<-x+k*t
xax<-xax+as.numeric(crossprod(x,avec))
xbx<-xbx+as.numeric(crossprod(x,bvec))
if (xbx<tol) stop("Geradin: xbx has become too small")
chcount<-n - length(which((xlast+offset)==(x+offset)))
if (trace) cat("Number of changed components = ",chcount,"\n")
pn<-xax/xbx # step 17 different order
if (chcount==0) {
  keepgoing<-FALSE # not really needed
  msg<-"Unchanged parameters -- done"
  break
}
if (pn >= p0) {
  if (trace) cat("RQ not reduced, restart\n")
  break # out of itn loop, not while loop (TEST!)
}
p0<-pn # step 19
g<-2*(avec-pn*bvec)/xbx
gg<-as.numeric(crossprod(g))
if (trace) cat("Itn", itn," RQ=",p0," after ",ipr," products, gg=",gg,"\n")
if (gg<tol){ # step 20
  if (trace) cat("Small gradient in iteration, restart\n")
  break # out of itn loop, not while loop (TEST!)
}
xbt<-as.numeric(crossprod(x,z)) # step 21
w<-y-pn*z # step 22
tabt<-as.numeric(crossprod(t,w))
beta<-as.numeric(crossprod(g,(w-xbt*g)))
beta<-beta/tabt # step 23
t<-beta*t-g
} # end loop on itn -- step 24
} # end main loop -- step 25
ans<-list(x=x, RQ=p0, ipr=ipr, msg=msg) # step 26
}

```

References

Nash, John C. 1979. *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation*. Bristol: Adam Hilger.