Sparse Matrices in package Matrix and applications

Martin Maechler and Douglas Bates

Seminar für Statistik ETH Zurich Switzerland

Department of Statistics
University of Madison, Wisconsin
(maechler|bates)@R-project.org (R-Core)

useR! 2009, Rennes July 10, 2009

Outline

- Introduction to Matrix and Sparse Matrices
 - Sparse Matrices in package Matrix
 - Matrix: Goals
 - 3D space of Matrix classes
- 2 Applications in Spatial Statistics
 - Regression with Spatially Dependent Errors: SAR(1)
- 3 Application Mixed Modelling (RE)ML in R
- 4 Who's the best liked prof at ETH?

• Matrix: the movie

Matrix: the R package

Package Matrix: a recommended R package since R 2

Infrastructure for other packages for several years, notably Ime4e

CRAN nowadays lists direct "reverse dependencies

- Matrix: the movie
- Matrix: the R package:
- Package Matrix: a recommended R package since R 2.9.0
- Infrastructure for other packages for several years, notably 1me4¹
- CRAN nowadays lists direct "reverse dependencies"

- Matrix: the movie
- Matrix: the R package:
- Package Matrix: a recommended R package since R 2.9.0
- ullet Infrastructure for other packages for several years, notably <code>lme4^1</code>
- CRAN nowadays lists direct "reverse dependencies":

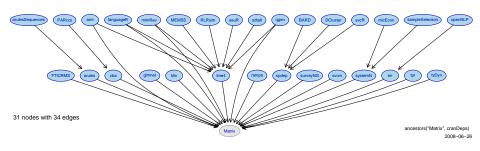
¹ Ime4 := (Generalized-) (Non-) Linear Mixed Effect Modelling,

(using S4 | re-implemented from scratch the 4th-time) (P + 4 P + 4

- Matrix: the movie
- Matrix: the R package:
- Package Matrix: a recommended R package since R 2.9.0
- ullet Infrastructure for other packages for several years, notably lme4 1
- CRAN nowadays lists direct "reverse dependencies":

(reverse) Dependencies on Matrix

On June 26, 2008 (> one year ago), Matrix was not yet recommended, and had the following CRAN dependency graph:



i.e., 14 + 1 directly dependent packages.

Dependencies on Matrix – 2009-07

Today, quite a few more packages depend on Matrix explicitly: ${\sf CRAN} \to {\sf Packages} \to {\sf Matrix} \quad {\sf displays} \; {\sf the} \; {\sf following} \\ {\sf http://cran.r-project.org/web/packages/Matrix/}$

Matrix: Sparse and Dense Matrix Classes and Methods

Classes and methods for dense and sparse matrices and operations on them using Lapack and SuiteSparse.

Version: 0.999375-29 Priority: recommended

Depends: R (\geq 2.9.0), stats, methods, utils, <u>lattice</u>

Imports: graphics, <u>lattice</u>, grid, stats

Enhances: graph, SparseM

Author: Douglas Bates and Martin Maechler

Reverse dependencies:

Reverse FAIR, FTICRMS, GOSim, MCMCglmm, Metabonomic, arm, arules, glmnet, klin, depends: languageR, lme4, mlmRev, pedigreemm, qgen, ramps, spdep, surveyNG, svcm,

systemfit, tpr, tsDyn

http://cran.r-project.org/web/packages/Matrix/:

Matrix: Sparse and Dense Matrix Classes and Methods

Classes and methods for dense and sparse matrices and operations on them using Lapack and SuiteSparse.

Version: 0.999375-29 Priority: recommended

Depends: $R (\geq 2.9.0)$, stats, methods, utils, lattice

Imports: graphics, lattice, grid, stats

Enhances: graph, SparseM

Author: **Douglas Bates and Martin Maechler**

Reverse dependencies:

Reverse FAIR, FTICRMS, GOSim, MCMCglmm, Metabonomic, arm, arules, glmnet, klin, depends:

languageR, Ime4, mlmRev, pedigreemm, qgen, ramps, spdep, surveyNG, svcm,

systemfit, tpr, tsDyn

Reverse imports:

arules, cba

R.matlab, RSiena, Rcsdp, blockmodeling, classGraph, e1071, gmodels, igraph, Reverse

suggests: rattle, spam, survey

Reverse

Rcplex, Rcsdp enhances:

useR!, Rennes 2009

Dependencies on Matrix — 2009-07 — Summary

- After one year, we have 22 (up from 15) packages depending on Matrix explicitly, plus another 12 "suggest" or "enhance" it.
- Ootably glmnet, Trevor Hastie's favorite in yesterday's keynote.
- Most important one: Ime4 and its dependencies

Dependencies on Matrix — 2009-07 — Summary

- After one year, we have 22 (up from 15) packages depending on Matrix explicitly, plus another 12 "suggest" or "enhance" it.
- Notably glmnet, Trevor Hastie's favorite in yesterday's keynote.
- Most important one: Ime4 and its dependencies

Intro to Sparse Matrices in R package Matrix

- The R Package Matrix contains dozens of matrix classes and hundreds of method definitions.
- Has sub-hierarchies of denseMatrix and sparseMatrix.
- Very basic intro in *some* of sparse matrices:

simple example — Triplet form

The most obvious way to store a sparse matrix is the so called "Triplet" form; (virtual class TsparseMatrix in Matrix):

```
> A < - spMatrix(10, 20, i = c(1,3:8),
+
     i = c(2,9,6:10),
     x = 7 * (1:7)
+
> A # a "dgTMatrix"
10 x 20 sparse Matrix of class "dgTMatrix"
```

Less didactical, slighly more recommended: A1 <- sparseMatrix(....)

simple example – 2 –

```
> str(A) # note that *internally* 0-based indices (i,j) are used
Formal class 'dgTMatrix' [package "Matrix"] with 6 slots
 ..0 i : int [1:7] 0 2 3 4 5 6 7
 ..0 j : int [1:7] 1 8 5 6 7 8 9
 ..@ Dim : int [1:2] 10 20
 ... @ Dimnames:List of 2
 ....$ : NULL
 .. ..$ : NULL
 ..0 x : num [1:7] 7 14 21 28 35 42 49
 ..@ factors : list()
> A[2:7, 12:20] \leftarrow rep(c(0,0,0,(3:1)*30,0), length = 6*9)
```

> A >= 20 ## <--- what result do you expect ?

simple example - 3 -

```
> A >= 20 # -> logical sparse; nice show() method
10 x 20 sparse Matrix of class "lgTMatrix"
[4,] . . . . . | . . . . . . . . . | | | . .
[5,] . . . . . . | . . . . | . . . . | | | .
[8,] . . . . . . . . | . . . . . . . . .
```

sparse compressed form

Triplet representation: easy for us humans, but can be both made smaller and more efficient for (column-access heavy) operations:

The "column compressed" sparse representation:

Column *index* slot j replaced by a column *pointer* slot p.

R Package Matrix: Compelling reasons for S4

- Classes for Matrices: well-defined inheritance hierarchies:
 - Content kind: Classes dMatrix, 1Matrix, nMatrix, (iMatrix, zMatrix) for contents of double, logical, pattern (and not yet integer and complex) Matrices, where nMatrix only stores the location of non-zero matrix entries (where as logical Matrices can also have NA entries)
 - sparsity: denseMatrix, sparseMatrix
 - 3 structure: general, triangular, symmetric, diagonal Matrices
- Inheritance: Visualisation via graphs
- Multiple Inheritance (of classes)
- Multiple Dispatch (of methods)

Multiple Dispatch in S4 for Matrix operations

Methods for "Matrix"-matrices: Often 2 matrices involved...

- ① x %*% y
- 3 tcrossprod(x,y) xy^{T}
- x + y "Arith" group methods

and many many more.

S4 >> S3

- S4 multiple dispatch: Find method according to classes of both (or more) arguments.
- S3 single dispatch: e.g., "ops.Matrix": only first argument counts.

Goals of Matrix package

- interface to LAPACK= state-of-the-art numerical linear algebra for dense matrices
 - making use of special structure for symmetric or triangular matrices (e.g. when solving linear systems)
 - setting and keep such properties allows more optimized code in these cases.
- Sparse matrices for large designs: regression, mixed models, etc
- …... [omitted in this talk]

Hence, quite a few different classes for matrices.

```
many Matrix classes . . .
> library(Matrix)
> length(allCl <- getClasses("package:Matrix"))</pre>
[1] 98
> ## Those called "...Matrix" :
> length(M.Cl <- grep("Matrix$",allCl, value = TRUE))</pre>
Γ17 70
i.e., many ..., each inheriting from root class "Matrix"
> str(subs <- showExtends(getClassDef("Matrix")@subclasses,</pre>
                            printTo=FALSE))
+
List of 2
 $ what: chr [1:76] "compMatrix" "triangularMatrix" "dMatrix" "iMatrix" ...
 $ how : chr [1:76] "directly" "directly" "directly" "directly" ...
> ## even more...: All those above and these in addition:
> subs$what[ ! (subs$what %in% M.Cl)]
[1] "Cholesky" "pCholesky" "BunchKaufman" "pBunchKaufman"
      a bit messy | . . .
                                              ▲□▶ ▲□▶ ▲□▶ ▲□▶ ● めぬぐ
```

3-way Partitioning of "Matrix space"

```
Logical organization of our Matrices: Three ( 3 ) main "class classifications" for our Matrices, i.e., three "orthogonal" partitions of "Matrix space", and every Matrix object's class corresponds to an intersection of these three partitions. i.e., in R 's S4 class system: We have three independent inheritence schemes for every Matrix, and each such Matrix class is simply defined to contain three virtual classes (one from each partitioning scheme), e.g,
```

3-way Partitioning of Matrix space — 2

The three partioning schemes are

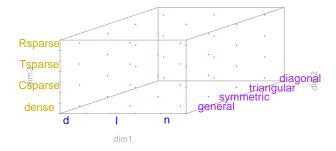
- Ontent type: Classes dMatrix, lMatrix, nMatrix, (iMatrix, zMatrix) for entries of type double, logical, pattern (and not yet integer and complex) Matrices.
 nMatrix only stores the location of non-zero matrix entries (where as logical Matrices can also have NA entries!)
- structure: general, triangular, symmetric, diagonal Matrices
- sparsity: denseMatrix, sparseMatrix

First two schemes: a slight generalization from LAPACK for dense matrices.

3D space of Matrix classes

three-way partitioning of Matrix classes visualized in 3D space, dropping the final Matrix, e.g., "d" instead of "dMatrix":

- > d1 <- c("d", "l", "n")
- > d2 <- c("general", "symmetric", "triangular", "diagonal")</pre>
- > d3 <- c("dense", c("Csparse", "Tsparse", "Rsparse"))</pre>
- > clGrid <- expand.grid(dim1 = d1, dim2 = d2, dim3 = d3, KEEP.OUT.A
- > clGr <- data.matrix(clGrid)</pre>
- > library(scatterplot3d)
 used for visualization:



3-fold classification — Matrix naming scheme

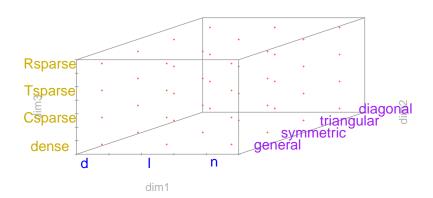
Actual classes follow a "simple" terse naming convention: > str(M3cl <- grep("^...Matrix\$",M.Cl, value = TRUE))</pre>

- "Actual" classes: Matrix objects are of those; the above "points in 3D space"
- 2 Virtual classes: e.g. the above coordinate axes categories. Superclasses of actual ones cannot have objects of, but —importantly— many *methods* for these virtual classes.

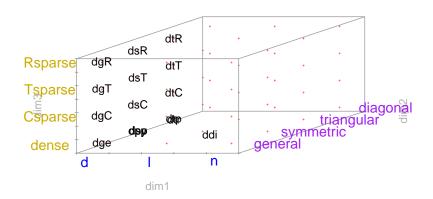
```
chr [1:47] "corMatrix" "ddiMatrix" "dgCMatrix" "dgeMatrix" ...
> substring(M3cl,1,3)
 [1] "cor" "ddi" "dgC" "dge" "dgR" "dgT" "dpo" "dpp" "dsC" "dsp" "dsR" "dsI
[13] "dsy" "dtC" "dtp" "dtr" "dtR" "dtT" "ldi" "lgC" "lge" "lgR" "lgT" "lsC
[25] "lsp" "lsR" "lsT" "lsy" "ltC" "ltp" "ltr" "ltR" "ltT" "ngC" "nge" "ngF
[37] "ngT" "nsC" "nsp" "nsR" "nsT" "nsy" "ntC" "ntp" "ntr" "ntR" "ntT"
```

> M3cl <- M3cl [M3cl != "corMatrix"] # corMatrix not desired in foll

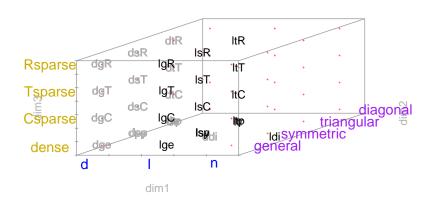
3D space of Matrix classes



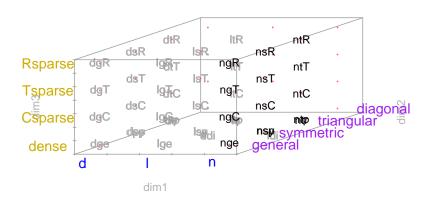
Matrix 3d space: filled (2)



Matrix 3d space: filled (3)



Matrix 3d space: filled (4)



Spatially Dependent Errors — SAR(1)

Regression with spatially dependent errors; observations at *locations* i, i = 1, ..., n, n in the thousands, possibly 100'000s.

Simultaneous Autoregression

$$y = X\beta + u$$
 where $u = \lambda Wu + \epsilon$. (1)

- W : matrix (W_{ij}) of "distance-based contiguities" of locations i and j $(W_{ii} \equiv 0)$.
- λ : SAR(1) parameter; estimate via MLE, (β profiled out).
- $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 (\boldsymbol{I} \lambda \mathbf{W})^{-1} (\boldsymbol{I} \lambda \mathbf{W}^{\intercal})^{-1})$
- For log likelihood, need to compute determinant $| \boldsymbol{I} \lambda \mathbf{W} | = (-\lambda)^n \left| -\mathbf{W} + \frac{1}{\lambda} \boldsymbol{I} \right|$ for many λ .

Compute Cholesky / Determinant of $A + \rho I$ for large sparse symmetric A: \implies Fast Cholesky Update

SAR(1) – fast Likelihood from Cholesky Update

```
Data provided by Roger Bivand, as a relevant test case:
```

```
> data(USCounties, package="Matrix")
> dim(USCounties)
[1] 3111 3111
> (n <- ncol(USCounties))</pre>
[1] 3111
> IM <- .symDiagonal(n)</pre>
> nWC <- -USCounties
> set.seed(1)
> rho <- sort(runif(50, 0, 1)) ## rho = 1 / lambda
and now compute determinant(A) =: |A|
```

$$|I - \lambda \mathbf{W}| \propto \left| -\mathbf{W} + \frac{1}{\lambda} I \right| \text{ for many } s.$$
 (2)

SAR(1) – Cholesky Update – 2 –

```
> ## Determinant : Direct Computation
> system.time(MJ <- sapply(rho, function(x)
         determinant(IM - x * USCounties, logarithm = TRUE)$modulus
  user system elapsed
 4.180 0.064 4.608
> ## Determinant : "high-level" Update of the Cholesky {Simplicial |
> C1 <- Cholesky(nWC, Imult = 2)
> system.time(MJ1 <- n * log(rho) +
    sapply(rho, function(x) c(determinant(update(C1, nWC, 1/x))$more
  user system elapsed
 0.656 0.000 0.715
> stopifnot(all.equal(MJ, MJ1))
> C2 <- Cholesky(nWC, super = TRUE, Imult = 2) ## <<-- "Supernodal"
> system.time(MJ2 <- n * log(rho) +
    sapply(rho, function(x) c(determinant(update(C2, nWC, 1/x))$more
```

◆ロト ◆個ト ◆量ト ◆量ト ■ める(で)

user system elapsed 0.772 0.036 0.867

SAR(1) - Cholesky Update - 3 > stopifnot(all.equal(MJ, MJ2)) > ## Determinant : "low-level" Update of the Cholesky {Simplicial / > system.time(MJ3 <- n*log(rho) + Matrix:::ldetL2up(C1, nWC,1/rho)) user system elapsed 0.400 0.000 0.415 > stopifnot(all.equal(MJ, MJ3)) > system.time(MJ4 <- n*log(rho) + Matrix:::ldetL2up(C2, nWC,1/rho)) user system elapsed</pre>

```
0.384 0.000 0.406
> stopifnot(all.equal(MJ, MJ4))
```

Findings:

- Using Cholesky update: order of magnitude faster
- 3 An even faster method for Det(Chol(.)) yields another 50% speed.

Mixed Modelling - (RE)ML Estimation in pure R

In (linear) mixed effects, the evaluation of the (RE) likelihood or equivalently deviance, needs repeated Cholesky decompositions of

$$\boldsymbol{U}_{\theta}\boldsymbol{U}_{\theta}^{\mathsf{T}} + \boldsymbol{I},\tag{3}$$

for many θ values (= the relative variance components) and (often very large), very sparse matrix U_{θ} where only the *non*-zeros of U depend on θ , i.e., the sparsity pattern is given (by the observational design). Sophisticated (fill-reducing) Cholesky done in two phases:

- $oldsymbol{0}$ "symbolic" decomposition: Determine the non-zero entries of $oldsymbol{L}$ $(oldsymbol{L}oldsymbol{L}^\intercal=oldsymbol{U}oldsymbol{U}^\intercal+oldsymbol{I}),$
- numeric phase: compute these entries.
 Phase 1: typically takes much longer; only needs to happen once.
 Phase 2: "update the Cholesky Factorization"

Who's the best liked prof at ETH?

- Private donation for encouraging excellent teaching at ETH
- Student union of ETH Zurich organizes survey to award prizes:
 Best lecturer of ETH, and of each of the 14 departments.
- Smart Web-interface for survey: Each student sees the names of his/her professors from the last 4 semesters and all the lectures that applied.
- ratings in $\{1, 2, 3, 4, 5\}$.
- high response rate

Who's the best prof — data

```
> md <- within(read.csv("~/R/MM/Pkg-ex/lme4/puma-lmertest.csv"), {</pre>
      s <- factor(s) # Student_ID
      d <- factor(d) # Lecturer_ID ("d"ozentIn)</pre>
+
+
      dept <- factor(dept)</pre>
      service <- factor(service)</pre>
+
+
      studage <- ordered(studage)## *ordered* factors</pre>
+
      lectage <- ordered(lectage) })</pre>
> str(md)
'data.frame': 73421 obs. of 7 variables:
$ s : Factor w/ 2972 levels "1", "2", "3", "4", ...: 1 1 1 1 2 2 3 3 3 3 .
         : Factor w/ 1128 levels "1", "6", "7", "8", ...: 525 560 832 1068 62 4
$ studage: Ord.factor w/ 4 levels "2"<"4"<"6"<"8": 1 1 1 1 1 1 1 1 1 1 1 ...
$ lectage: Ord.factor w/ 6 levels "1"<"2"<"3"<"4"<...: 2 1 2 2 1 1 1 1 1 1</pre>
$ service: Factor w/ 2 levels "0","1": 1 2 1 2 1 1 2 1 1 1 ...
$ dept : Factor w/ 15 levels "1","2","3","4",..: 15 5 15 12 2 2 14 3 3 3
$ y : int 5 2 5 3 2 4 4 5 5 4 ...
```

Modelling the ETH teacher ratings

Model: The rating depends on

- students (s) (rating subjectively)
- teacher (d) main interest
- department (dept)
- "service" lecture or "own department student", (service: 0/1).
- semester of student at time of rating (studage $\in \{2,4,6,8\}$).
- how many semesters back was the lecture (lectage).

Main question: Who's the best prof?

Hence, for "political" reasons, want d as a **fixed** effect.

Model for ETH teacher ratings

Want d ("teacher_ID", ≈ 1000 levels) as **fixed** effect. Consequently, in

$$y = X\beta + Zb + \epsilon$$

have \boldsymbol{X} as $n \times 1000$ (roughly), \boldsymbol{Z} as $n \times 5000$, $n \approx 70'000$.

sparseX = TRUE: sparse X (fixed effects) in addition to the indispensably sparse Z (random effects).

Unfortunately: Here, the above "sparseX - Imer" ends in

Error ... Cholmod error 'not positive definite' at file:../Cholesky/.....

Good News: Newly in Matrix:

sparse.model.matrix()

- which lmer() can use,
- or you can use for "truly sparse" least squares (i.e. no intermediately dense design matrix)
- something we plan to provide in Matrix 1.0-0.

- Recommended R package "Matrix"
- Sparse Matrices: in increasing number of applications
- S4 classes and methods are the natural implementation tools
- Ime4 is going to contain an alternative "pure R" version of ML and REML, you can pass to nlminb() (or optim() if you must :-).
 UseRs can easily extend these R functions to more flexible models or algorithms.
- Matrix 1.0-0
 - will happen
 - O will contain sparse.model.matrix()
 - → will contain truly sparse lm(*, sparse=1RUE)

- Recommended R package "Matrix"
- Sparse Matrices: in increasing number of applications
- S4 classes and methods are the natural implementation tools
- Ime4 is going to contain an alternative "pure R" version of ML and REML, you can pass to nlminb() (or optim() if you must :-).
 UseRs can easily extend these R functions to more flexible models or algorithms.
- Matrix 1.0-0
 - will contain sparse wodel metric (
 - will contain sparse moder matrix (
 - will contain truly sparse lm(*, sparse=THUE)

- Recommended R package "Matrix"
- Sparse Matrices: in increasing number of applications
- S4 classes and methods are the natural implementation tools
- Ime4 is going to contain an alternative "pure R" version of ML and REML, you can pass to nlminb() (or optim() if you must :-).
 UseRs can easily extend these R functions to more flexible models or algorithms.
- Matrix 1.0-0

- Recommended R package "Matrix"
- Sparse Matrices: in increasing number of applications
- S4 classes and methods are the natural implementation tools
- Ime4 is going to contain an alternative "pure R" version of ML and REML, you can pass to nlminb() (or optim() if you must :-).
 UseRs can easily extend these R functions to more flexible models or algorithms.
- Matrix 1.0-0

- Recommended R package "Matrix"
- Sparse Matrices: in increasing number of applications
- S4 classes and methods are the natural implementation tools
- Ime4 is going to contain an alternative "pure R" version of ML and REML, you can pass to nlminb() (or optim() if you must :-).
 UseRs can easily extend these R functions to more flexible models or algorithms.
- Matrix 1.0-0
 - will happen
 - ② will contain sparse.model.matrix()
 - will contain truly sparse lm(*, sparse=TRUE)

- Recommended R package "Matrix"
- Sparse Matrices: in increasing number of applications
- S4 classes and methods are the natural implementation tools
- Ime4 is going to contain an alternative "pure R" version of ML and REML, you can pass to nlminb() (or optim() if you must :-).
 UseRs can easily extend these R functions to more flexible models or algorithms.
- Matrix 1.0-0
 - will happen
 - @ will contain sparse.model.matrix()
 - will contain truly sparse lm(*, sparse=TRUE)

- Recommended R package "Matrix"
- Sparse Matrices: in increasing number of applications
- S4 classes and methods are the natural implementation tools
- Ime4 is going to contain an alternative "pure R" version of ML and REML, you can pass to nlminb() (or optim() if you must :-).
 UseRs can easily extend these R functions to more flexible models or algorithms.
- Matrix 1.0-0
 - will happen
 - will contain sparse.model.matrix()
 - will contain truly sparse lm(*, sparse=TRUE)

- Recommended R package "Matrix"
- Sparse Matrices: in increasing number of applications
- S4 classes and methods are the natural implementation tools
- Ime4 is going to contain an alternative "pure R" version of ML and REML, you can pass to nlminb() (or optim() if you must :-).
 UseRs can easily extend these R functions to more flexible models or algorithms.
- Matrix 1.0-0
 - will happen
 - will contain sparse.model.matrix()
 - will contain truly sparse lm(*, sparse=TRUE)

- Recommended R package "Matrix"
- Sparse Matrices: in increasing number of applications
- S4 classes and methods are the natural implementation tools
- Ime4 is going to contain an alternative "pure R" version of ML and REML, you can pass to nlminb() (or optim() if you must :-).
 UseRs can easily extend these R functions to more flexible models or algorithms.
- Matrix 1.0-0
 - will happen
 - will contain sparse.model.matrix()
 - will contain truly sparse lm(*, sparse=TRUE)

- Recommended R package "Matrix"
- Sparse Matrices: in increasing number of applications
- S4 classes and methods are the natural implementation tools
- Ime4 is going to contain an alternative "pure R" version of ML and REML, you can pass to nlminb() (or optim() if you must :-).
 UseRs can easily extend these R functions to more flexible models or algorithms.
- Matrix 1.0-0
 - will happen
 - will contain sparse.model.matrix()
 - will contain truly sparse lm(*, sparse=TRUE)