# Working Document for the Improvement to nls() GSOC project

Arkajyoti Bhattacharjee, Indian Institute of Technology, Kanpur
John C. Nash, University of Ottawa, Canada
Heather Turner, University of Warwick, UK

02 July, 2021

## TODOS

- Show how `dir` works in nlsalt::numericDeriv(); need some examples/tests Currently is used when approaching upper bounds.

## Abstract

nls() is the primary nonlinear modeling tool in base R. It has a great many features, but it is about two decades old and has a number of weaknesses, as well as some gaps in documentation. This document is an ongoing record of work under the Google Summer of Code 2021 of the first author. As such it is NOT meant to be a finished academic report, but a form of extended diary of activity, issues and results.

## Project repository

https://gitlab.com/nashjc/improvenls

## ——— DIARY PORTION ———

- In reverse date order *

### [2021-7-1] to [2021-7-?] nlsj

The following is from an email to Patrice Kiener, a co-mentor on the Google Summer of code NNBenchmark projects for 2019 and 2020.

```
The nls() stuff is going much more slowly than I'd like. I knew when we started that
the student is willing and generally capable but not very experienced in R internals.
But then, my own experience is strong on nonlinear optimization, but not so much on
the quirkiness of R + C interaction. Heather and I decided to try to segment the work
so there are some sub-projects that can be completed, though they might not be
directly on the original topic. I'm leading the code digging, and getting Arkajyoti
to try to follow and ask questions to force us to keep honest. This is working quite
well, and between us we are learning and moving the frontier forward. However, as my
doctoral supervisor (C. A. Coulson FRS) said "pi is the ratio of how long a task
takes over how long you thought it would take".

The big obstacle at the moment is a key routine created by nls() in its "model"
object "m" called "setPars". Here is the output of fit <- nls([a simple problem])


fit$m$setPars
```

```
function(newPars) {
         setPars(newPars)
         resid <<- .swts * (lhs - (rhs <<- getRHS())) # envir = thisEnv {2 x}
         dev   <<- sum(resid^2) # envir = thisEnv
         if(length(gr <- attr(rhs, "gradient")) == 1L) gr <- c(gr)
         QR <<- qr(.swts * gr) # envir = thisEnv
         (QR$rank < min(dim(QR$qr))) # to catch the singular gradient matrix
         }
<bytecode: 0x55f3e6c05cb0>
<environment: 0x55f3e6c27360>
>

You'll note:
 1) it is recursive
 2) it creates a bunch of functions and objects, but only returns a single LOGICAL
 3) if you run nls() finding QR and other objects just doesn't seem possible easily
   (I haven't been able to find them! But they are the key solution item.)
 4) "global" assignments are used "<<-"
 5) "hidden" objects are used (.swts)
 6) if the problem has nonlinear parameters in the lhs, then this fails, because it
    only builds the jacobian for the rhs of the model. Problems like
    that aren't traditional in statistics, but can arise in optimization.

So a real mess. I've been in contact with Doug Bates, who is supposedly the author and who
said he would help if he could. He switched to Julia a few years ago. He can't recall why
the recursion is needed. But if you run the function outside of nls(), it blows the C
language stack! Sigh.

I plan to refactor, eliminating the recursion and globals and use a proper list() return
so all objects are in known locations. However, that will certainly cause us some pain when
it comes to profile.nls() and possibly some other features. And truthfully, I'd like to
understand the code, even if we replace it. My intention is to continue for a few more
days to run calls of the function under different initial conditions to create a
"reproducible example" (I think reprex is becoming the abbreviation) so I can post on
r-devel and see if there is anyone who can help to decode this. I strongly suspect that
the R-core maintainers are praying that no changes in the R and C infrastructure will
break the current code, as I'm pretty sure there's no developer documentation to provide
a guide.

On the bright side, I have got numericDeriv() function all in R. It's a bit slower than the
mixed R+C, but much shorter and more easily maintained and adapted -- the plan is to offer
the analytic derivatives as an option. Also to change the name to "jacobian" or "jacobian.nls".
And we've drafts of several short articles on some issues tangential to the project,
including one I'm preparing on the variety of ways nonlinear least squares can be
structured and implemented. Also one on a neat idea from Duncan Murdoch on how to extract
base-R code into a package that can be modified, thereby avoiding rebuilds of full R.
```

In that JN has NOT been able to satisfactorily understand the structure of setPars() and setVarying() (the latter produces a huge set of functions, and it is not clear how it has been defined), I decided to do a zero-base build of a new set of functions in a package called `nlsj`. This will hopefully have fully documented functions and features. However, I expect that it will need quite a lot of work to add the equivalent of `profile.nls()` and similar features. We note, from the `nls()` documentation, that

```
An nls object is a type of fitted model object. It has methods for the generic
functions anova, coef, confint, deviance, df.residual, fitted, formula, logLik,
```

`predict, print, profile, residuals, summary, vcov and weights.`

Thus we will have to build equivalent capability, but we can possibly avoid a lot of the extraneous material in the current infrastructure and also document what we do provide.

**[2021-6-29] and [2021-6-30] Struggling with nls() structure**

- JN: worked on the Croucher-expandednlsnoc.R file. A number of issues related to the setPars() function. This was expanded to try to work around the issue that it is recursively defined (why?). Then troubles with getRHS(), which does not seem to be exposed for access in the top level code. Moreover, issues with .swts, which is hidden (why? again).
- JN: discussion with Paul Gilbert about recursive calls and global assignment in nlsModel and setPars. Email to Doug Bates got rather discouraging reply that Doug unable to recall or figure out why setPars is recursive.
- setPars(): search shows it is in nls-profile() as well as nls(). The calls to it are unusual, first with no argument, then with an argument. We need to understand how it works, though nls-profile is a tool that may not be sensible to keep in R. However, "profile" is a tool in R, and "profile.nls" is documented, with an example.
- note also that setPars() works with gradient in rhs only. This is NOT OK when lhs and rhs have parameters, which can be the case when we are not modeling a single variable. Should use `resid`.

**[2021-6-28] continuing Croucher-expanded**

- JN: found that minpack.lm version of conv could be put into code and appears to function correctly. It is much simpler than code in nlspkg::nlsModel(). Still need to sort out setPars.
- Online meeting – some issues with git, Google Meet. Discussed diary portion of this document, some issues related to testing, and plans for this week. Agreed to meet Sunday July 4 to avoid collision with UseR!2021.

**[2021-6-26] and [2021-6-27]**

- expanded VarietyInNonlinearLeastSquaresCodes substantially
- started unrolling nls code in Croucher examples. Some issues with if ... else and brackets, possibly due to last "}" in function definition of nls().
- Note the strange "%||%" – comes from `utils` package via base-internals.R in `nlspkg` or `nlsalt` BUT looks like a language construct. ??Should we replace with something inline so this does not trip up other programmers. In `minpack.lm` the structure used is

```
env <- environment(formula)
if (is.null(env)) env <- parent.frame()
```

- AB: looking up suitable examples in NRAIA package[23 June]
- AB: looking up various testing methods in R [21-28 June]
- AB: working on an Rmd file, documenting what testing strategies might be taken up for our project [26-27 June] (Made a branch, but I don't see a new branch created in the repo. May have forgotten how to do this. Will look into it.

**[2021-6-23] [2021-6-25] Tidying and documenting nlsModel**

- JN cleaned up DerivsNLS. Extracted ssasymptot model and tested in nlsalt. numericDeriv must be working, but not explicitly tested. Note that the encapsulated version in the examples of nls() seem to fail.??

- used file inport of nls.R and nls.c parts (which are now separated out in improvnls)
- nlsModelDoc proceeding slowly. There are many undocumented features and pieces of code in nlsModel() that need careful understanding, particularly the scoping assignment "«-" in some lines of the code. In `minpack.lm` there is a version of nlsModel() that uses `assign` rather than `<<-`, and some documentation points out that this may be a more explicit mechanism (and therefore easier to understand and keep correct). Note the R documentation `?assignOps` as well as the `assign()` and `get()` functions.
- created nlsModelx.R file with nlsModelx(), and then added print statements to allow understanding of the function and build developer documentation.
- AB: looking up some tutorials on NLS with the aim of writing an article/tutorial on NLS, with the use of nls_new[24-25 June]

**[2021-6-22] Answering some of the TODOS / nls-noc.R / nlsModelDoc.Rmd**

- Do we need to update nlsalt and nlspkg separately from improvenls on a machine when there are changes imported via git? Answer is that git within improvenls directory seems sufficient.
- response from Doug Bates. Added him to gitlab project.
- Expanded document nlsModelDoc.Rmd to show what different objects created actually do. There are still plenty of questions, especially about `setPars()`.
- AB: in our documents, I am currently copying and pasting material. However, it would be sensible to do an "include" of the material from the nlsalt package, but to do that cleanly, we will want to separate the functions into different files. This will mean we have to be careful that the package still builds OK.

  This can be accomplished via (following commented in the Rmd file). Note that the issue is addressed in several ways in `/home/gramps/GrampsSetup/GrampsSetup.Rmd`, which needs to be better explained here.

https://towardsdatascience.com/rtest-pretty-testing-of-r-packages-50f50b135650

https://towardsdatascience.com/unit-testing-in-r-68ab9cc8d211

https://r-pkgs.org/tests.html

JN contacted Doug Bates who indicated he may be able to provide some input on the code. There does not appear to be any developer documentations. Ongoing work with nlsModel() seems to indicate that the "m" object is set up to provide all the infrastructure for the rest of the code to work with for a particular nonlinear model.

- AB: reading up parts of chapter 2 and 3 of Bates and co's NRAIA book[21-22 June]

**Week of [2021-6-21]**

Online meeting on June 21 talked about `nls-noc.R` as well as testing schemes. AB will try to develop a short appraisal of testing schemes and suggest a path for us. Some discussion of tools.

JN started `nlsModelDoc.Rmd`. There are many functions, seemingly without documentation. Also quite complex codes and reuse of names. In some codes the "«-" assignment is used. We will need to be very careful to document and understand this.

**[2021-6-21] Meeting etc.**

Met online.

- discussed testing schemes. AB will see if he can develop a short appraisal and suggestions for our use.
- talked about `nls-noc.R` process. This is ongoing by JN

- later in day, JN contacted Doug Bates to ask if there are developer notes or other information, as there seem to be extra features in the nlsModel() output and the nls_iter code in nls.c

**[2021-6-15] - [2021-6-20]: nlsalt issues**

1) Need to ensure the name "nlsalt" is fixed DESCRIPTION, NAMESPACE (dynload line), .gitignore (.so line can be *.so), init.c (dynload line, R_init line)

2) appears new all-R numericDeriv not quite right for all nls examples. Probably (to be determined??) whether example uses a function or expression in the definition of the model.

During this time did some work on splitting nls.R into separate files where major functions were in their own file. In particular, nlsModel() was separated out so that nlsModel and nlsModel.plinear did not get confused in editing, since they are largely the same.

The file `nls-noc.R` has the ongoing work to try to build an all-R nls_iter function. This is gradually proceeding, but not functional at the time of our June 21 online meeting.

JN took a brief look at 2010 UseR! tutorial on nonlinear models. This may give us some ideas for documentation and tests.

**[2021-6-12] –> [2021-6-14]**

Document PkgFromRbase written to document the packaging effort and report on the Makevars.win issue.

JN diverted to update optimx package.

**[2021-6-15] chunk code from file**

The **R Markdown Cookbook** is quite useful in giving hints on some features of R Markdown. In particular, sections 16.1 to 16.3 of

https://bookdown.org/yihui/rmarkdown-cookbook/

show how to use R scripts saved in files within chunks of R Markdown documents. This is important, as we may wish to list such files, but copying the script into the text of the document means that this copy and the original file can become unsynchronized.

From work on the Gramps genealogy software (and some local setup issues), JN found that the following chunk will allow listings in a nice format. Note the "comment=NA" to avoid line prefixing with hash symbols. The code which follows needs surrounding by the usual three ticks before and after.

```
{r lscript, comment=NA, echo=FALSE}
cat(readLines('/home/gramps/GrampsSetup/FrohnGramps'), sep = '\n')
```

**[2021-6-6] –> [2021-6-11]**

Had online meeting and got Rstudio more or less set up to deal with version control and package building, BUT ...

hit a big problem with 32 bit Windows architecture NOT linking to BLAS. A posting to R-package-devel list got a suggestion from Dirk Eddelbeutel to check sessionInfo() and this shows that on the Windows R side, BLAS and LAPACK are not listed, while they are in Linux. Duncan Murdoch pointed to makevars.win, but severeal tries with some possible PKG_LIBS settings did not give any change. Noted that command line CMD.exe use of "R CMD build" and "R CMD INSTALL" showed the issue to be at the INSTALL stage, and it is clear that the 32 bit architecture is NOT finding the BLAS routines, but

Then found https://stackoverflow.com/questions/42118561/error-in-r-cmd-shlib-compiling-c-code and this shows what to use for makevars.win, which is the single line file

PKG_LIBS = $(LAPACK_LIBS) $(BLAS_LIBS) $(FLIBS)

**[2021-6-6]**

Wanted package benchmarkme.

install failed. Needed dependencies.

So in linux, install libssl.dev libcurl4-openssl-dev

Then benchmarkme installed OK.

We can use this within our benchmarking in addition to sessionInfo().

**[2021-6-5]**

- created script `tnlsModel.R` as first step to create and produce an "m" object
- Current understanding (JN): nlsModel (and nlsModel.plinear too probably) creates an R object that it labels as class "nlsModel." This object contains functions that are called from nls.c::nls_iter to run the interation and estimate the parameters in the model. There seem to be some extraneous functions, and we can hopefully learn enough to remove the extras.
- The current structure is to particularize the functions in "m" so the (essentially) external nls.c code acts on these. As a first goal, and part of learning how things work, we will want to replace the nls.c::nls_iter with all-R equivalent.
- Started document nlsModel.Rmd to document this function. ?? AB: in our documents, I am currently copying and pasting material. However, it would be sensible to do an "include" of the material from the nlsalt package, but to do that cleanly, we will want to separate the functions into different files. This will mean we have to be careful that the package still builds OK.

## From `nlsr` vignette `nlsr-devdoc.Rmd` [2021-5-23 onwards]

**[2021-6-4]**

- Code for `numericDeriv()` in nls.R was extracted to `numericDeriv20210603CommentedWithC.txt`.
- The code for numeric_deriv was removed from nls.c
- The line in init.c that mentions numeric_deriv was commented out so it would not be linked into the shared library for nlsalt (nlsalt.so)
- By launching Rstudio from the nlsalt.Rproj file, the package `nlsalt` could be rebuilt using "clean and rebuild." nlsalt.Rproj can be initiated from the "Files" menu of Rstudio. Note that rebuild is probably necessary on each machine on which nlsalt is used. (?? Do we need to check this is the case?)

## [2021-6-2] Converting codes to All-R – numericDeriv

In the `ExDerivs.R` script, JN tested a forward difference code in R against a call to numericDeriv(). On a single example, it tests OK, so next step is to modify nls.R in package `nlsalt`.

Note that `ExDerivs.R` is in flux. "Old" tests will be commented out. Look for the double question mark (??) for things that may need to be checked.

Some concerns:

- the code requires us to get() and assign() values in the environment `rho` that is used to evaluate the nonlinear residual functions. There may or may not be ways to make this more efficient.
- the C code uses explicit loops. A very preliminarly use of R's vectorization in the difference between two residual vectors (i.e., without subscripts or looping) worked fine, but there are almost certainly further optimizations. By contrast, we MAY want to leave either explicit loops or commented code that contains such loops as a way to provide explanation of what is happening.

The code from the trials in ExDerivs.R was ported to nls.R in nlsalt package and tested with TestnumericDeriv1.R (added to repo). Things seem to work OK. I have NOT tried changing `dir`. ??

## [2021-6-1] nlsalt package mods started

numericDeriv first cut at pure R. JN incorporated C code into the R code of the numericDeriv function as comments, then tried to provide R statements to do what it appears the C is doing. Some steps and checks likely left out.

## [2021-6-1] nlspkg package created

Thanks to Duncan Murdoch who created nlspkg from the R source files. This is a package version of the files in the source tree that will let us work with the code and change it. When we issue

```
library(nlspkg)
```

once we have installed it, the objects in the package will replace (mask) those in base-R.

To build nlspkg, I needed the inconsolata.sty font that is in the `texlive-fonts-extra` package of Linux Mint. In other systems, this font may reside elsewhere.

As of June 1, the package builds and mostly checks. There are two warnings:

- a number of function objects are NOT documented. We may or may not wish to add `man` files (of type `.Rd`). The list of objects is quite long, but they seem not to be needed by the user. Note that nlminb is present, likely for the PORT option of the "algorithm" parameter of `nls()`.
- There is a non-standard license specification. Checking in the Writing R Extensions manual (file r-exts.pdf), suggests dropping one of the two licenses listed (GPL-2 and GPL-3). Dropping the latter removed one warning.

The package `nlspkg` is to provide the EXISTING nls() functionality, while the package `nlsalt` is for our replacement. I plan to make this all-R. If we want to use C (or something else) later, then we will create another "alternate" package.

Note that each of these packages needs its own "project." In Rstudio we choose to create a project using an existing directory (copy of that for `nlspkg`). This sets up the control files for the project and allows the package to be built and checked with the Rstudio tools. The umbrella ImproveNLS project handles the **git** version control.

### [2021-5-31] Notes on investigations

1) AB managed to get R built from an R-devel expanded tarball. JN had managed this earlier. There appear to be some transient glitches (HT has noted also). These are likely due to R-core members making trial changes. MASS package in particular noted.

2) Investigations of using subversion or git repo of the code so we can track any changes in files relevant to nls(). JN tried from https://github.com/wch/r-source/wiki which has some instructions. These seemed to work, but we note the need for texinfo and git-svn packages to be installed. Also standard "git pull" cannot be used. "git sv rebase" followed by "git svn fetch" tried and seem to work, but JN does not feel he understands what is going on.

3) Started to look into development of replacement code for parts of nls(). Created ReplaceNLS.R and RestoreNLS.R with the intention that source() of these files would replace the functions. However, first attempt, which simply added a cat() output to numericDeriv() function got

```
Error in numericDeriv(rexpr, theta, rho = weedenv) :
  object 'C_numeric_deriv' not found
```

## [JN: 2021-5-19] Some thoughts on how nls() code might be modified

- R-core will veto any changes that do not leave existing tests and examples as they are. Therefore we will be always thinking of ways to leave the default behaviour as it is now. In the relative offset convergence test that was causing small-residual problems to fail, JN suggested a small change to the test. In very simplified terms, the test is a ratio of `(new sum of squares)/(baseline sum of squares)` When both these quantities are small, we risk zero/zero situation. Therefore, a new control parameter `convTeestAdd` that has a default of 0 was introduced. This can be set in the `nls.control()` function. It is added to the denominator and allows the `nls()` function to complete satisfactorily when we have problems where the fit is nearly exact. Thus we are seeking similar modifications that are invoked by setting new control parameters to non-default settings to get new behaviour.
- In the material below on implementing nonlinear least squares, it should be possible to have a zero valued default for the Marquardt parameter $\lambda$. As long as this zero value is preserved, the legacy behaviour is used. Our decisions need to be careful in situations where we really do need the Marquardt stabilization. That is, should we simply fail (legacy behavior), should we override the legacy behaviour, since the user almost certainly does not want an avoidable failure, or should we seek an additional control that allows for more nuanced outcomes. Clearly thought is needed, and we will need to find some examples to illustrate our arguments.
- Note that nls() is a mix of several codes, using different solution tools. We will need to decide where our modifications apply and document carefully. JN opinion: we may want to postpone consideration of "port" and "plinear" algorithms, but it would be useful to document what they do. The current documentation suggests that bounds on parameters can only be used with "port."
- The derivative code to generate the Jacobian in nls() is (JN believes??) a simple forward-difference code. This may be in C rather than R. It is well-known that central differences, while they cost twice the computing effort, do a better job. Can we find a way to allow that easily?
- If there are bounds on parameters, the "step" taken in estimating differences may violate bounds. This is an issue well-known to workers in optimization, but not on the radar of most users. Clearly a "nasty." Can we do anything to at least warn users in a way that is useful?

## 2021-5-18

By email, a Google Meet was set up for May 19 at noon Ottawa time. JN downloaded the latest version of `R-devel.tar.gz` and unpacked it into a directory `~/vmshare/R-devel` which is shared with a VirtualBox VM of Kubuntu 20.04 (Focal Fossa), a long-term support version of Linux. Opening a terminal inside this VM, it was possible to build and run this version of R.

```
cd /media/sf_vmshare/R-devel/
./configure
make
sudo make install
[admin password]
R
```

This launched the development version of R correctly.

JN also set up this document.

### Agenda for May 19

- check Meet is working
- introductions
- start linux VM install if an iso is available. May want to check that VirtualBox Guest additions is available and that the shared directory works, as these sometimes require some attention to permissions and ownership etc.

- Consider early goals to for possible nls() changes. See below.
- Set objectives for next two weeks
- Set next online meeting

**Possible early goals**

- Get a VM running under VirtualBox and install build tools. See https://support.rstudio.com/hc/en-us/articles/218004217-Building-R-from-source I did NOT need more than ./configure in my build, as I am not building a server version.
- Try the build. (Cheer loudly when it works!)
- Explore and document the R source for nls-related code. In particular, we want
  - to list the files that have such code or calls to it
  - to note, if possible, what each does
  - to note, in particular, where nls() solves the Gauss-Newton equations, as we will want to modify these sections to augment them to allow a Marquardt stabilization.
  - to note, in particular, where nls() computes the Jacobian and/or Hessian for the nonlinear least squares problem. In package `nlsr`, there are tools that allow an expression for the model to be parsed and processed to compute the Jacobian using symbolic or automatic derivatives. This may or may not be feasible for `nls()`. However, we may be able to improve the approximation used.

    [2021-5-27] first draft DerivsNLS.Rmd. Note discovered "central" control for numericDeriv() function. It does NOT appear that the 'dir' control is actually used. This would likely change from forward to backward differences for SOME components of the step to compute first order finite difference approximation to the Jacobian.
- Augment this document to record what has been done and results or problems.
- Add to the bibliography file specified. This is taken from another work, but as Rmarkdown only uses the references it needs, it will serve as a template.
- AB can ask for pointers to references to add to this document or to subsidiary documents we will create as necessary to describe parts of the work if required.

[2021-5-25] Meet session. AB got Linux Mint 20.1 VM going and verified. SOme questions sorted out. JN added some material re: linear least squares solutions to the Variety document.

==========end diary only portion ==========

**5. Implementation of nonlinear least squares methods**

This section is a review of approaches to solving the nonlinear least squares problem that underlies nonlinear regression modelling. In particular, we look at using a QR decomposition for the Levenberg-Marquardt stabilization of the solution of the Gauss-Newton equations.

**Gauss-Newton variants**

Nonlinear least squares methods are mostly founded on some or other variant of the Gauss-Newton algorithm. The function we wish to minimize is the sum of squares of the (nonlinear) residuals $r(x)$ where there are $m$ observations (elements of $r$) and $n$ parameters $x$. Hence the function is

$$f(x) = \sum_i ({r_i}^2)$$

Newton's method starts with an original set of parameters $x[0]$. At a given iteration, which could be the first, we want to solve

$$x[k+1] = x[k] - H^{-1}g$$

where $H$ is the Hessian and $g$ is the gradient at $x[k]$. We can rewrite this as a solution, at each iteration, of

$$H\delta = -g$$

with

$$x[k+1] = x[k] + \delta$$

For the particular sum of squares, the gradient is

$$g(x) = 2 * r[k]$$

and

$$H(x) = 2(J'J + \sum_i (r_i * Z_i))$$

where $J$ is the Jacobian (first derivatives of $r$ w.r.t. $x$) and $Z_i$ is the tensor of second derivatives of $r_i$ w.r.t. $x$). Note that $J'$ is the transpose of $J$.

The primary simplification of the Gauss-Newton method is to assume that the second term above is negligible. As there is a common factor of 2 on each side of the Newton iteration after the simplification of the Hessian, the Gauss-Newton iteration equation is

$$(J'J)\delta = -J'r$$

This iteration frequently fails. The approximation of the Hessian by the Jacobian inner-product is one reason, but there is also the possibility that the sum of squares function is not "quadratic" enough that the unit step reduces it. Hartley (1961) introduced a line search along delta, while Marquardt (1963) suggested replacing $J'J$ with $(J'J + \lambda * D)$ where $D$ is a diagonal matrix intended to partially approximate the omitted portion of the Hessian.

Marquardt suggested $D = I$ (a unit matrix) or $D =$ (diagonal part of $J'J$). The former approach, when $\lambda$ is large enough that the iteration is essentially

$$\delta = -g/\lambda$$

gives a version of the steepest descents algorithm. Using the diagonal of $J'J$, we have a scaled version of this (see https://en.wikipedia.org/wiki/Levenberg%E2%80%93Marquardt_algorithm; Levenberg (1944) predated Marquardt, but the latter seems to have done the practical work that brought the approach to general attention.)

Nash (1977) found that on low precision machines, it was common for diagonal elements of $J'J$ to underflow. A very small modification to solve

$$(J'J + \lambda(D + \phi I))\delta = -g$$

where $\phi$ is a small number avoids most of these conditions. $\phi = 1$ seems to work quite well. We note that this modification would likely not have been recognized had I not been working on machines with mediochre floating-point arithmetic – a little more than 6 decimal digits of precision and no extended precision.

**Choices**

Both `nlsr::nlxb()` and `minpack.lm::nlsLM` use a Levenberg-Marquardt stabilization of the iteration described above, with `nlsr` using the modification involving the $\phi$ control parameter. The complexities of the code in `minpack.lm` are such that I have relied largely on the documentation to judge how the iteration is accomplished. `nls()` uses a straightforward Gauss-Newton iteration, but rather than form the sum of squares and cross-products, uses a QR decomposition of the matrix $J$ that has been found by a forward difference approximation.

(Nash (1979)), solving

$$(J^T J + \lambda D)\delta = -J^T r$$

where $D$ is some diagonal matrix and lambda is a number of modest size initially. Clearly for $\lambda = 0$ we have a Gauss-Newton method. Typically, the sum of squares of the residuals calculated at the "new" set of parameters is used as a criterion for keeping those parameter values. If so, the size of $\lambda$ is reduced. If not, we increase the size of $\lambda$ and compute a new $\delta$. Note that a new $J$, the expensive step in each iteration, is NOT required.

As for Gauss-Newton methods, the details of how to start, adjust and terminate the iteration lead to many variants, increased by different possibilities for specifying $D$. See Nash (1979).

**Using matrix decompositions [2021-5-23 onwards]**

In Nash (1979), the iteration equation was solved as stated. However, this involves forming the sum of squares and cross products of $J$, a process that loses some numerical precision. A better way to solve the linear equations is to apply the QR decomposition to the matrix $J$ itself. However, we still need to incorporate the $\lambda I$ or $\lambda D$ adjustments. This is done by adding rows to $J$ that are the square roots of the "pieces." We add 1 row for each diagonal element of $I$ and each diagonal element of $D$.

In each iteration, we reduce the $\lambda$ parameter before solution. If the resulting sum of squares is not reduced, $\lambda$ is increased, otherwise we move to the next iteration. Various authors (including the present one) have suggested different strategies for this. My current opinion is that a "quick" increase, say a factor of 10, and a "slow" decrease, say a factor of 0.2, work quite well. However, it is important to check that $\lambda$ has not got too small or underflowed before applying the increase factor. On the other hand, it is useful to be able to set lambda = 0 in the code so that a pure Gauss-Newton method can be evaluated with the program(s). The current code `nlfb()` uses the line

```
if (lamda<1000*.Machine$double.eps) lamda<-1000*.Machine$double.eps
```

to ensure we get an increase. To force a Gauss-Newton algorithm, the controls `laminc` and `lamdec` are set to 0.

**The Levenberg-Marquardt adjustment to the Gauss-Newton approach is the second major improvement of `nlsr` (and also its predecessor `nlmrt` and the package `minpack-lm`) over `nls()`.**

`nls()` attempts to minimize a sum of squared residuals by a Gauss-Newton method. If we compute a Jacobian matrix $J$ and a vector of residuals $r$ from a vector of parameters $x$, then we can define a linearized problem

$$J^T J\delta = -J^T r$$

This leads to an iteration where, from a set of starting parameters $x_0$, we compute

$$x_{i+1} = x_i + \delta$$

This is commonly modified to use a step factor `step`

$$x_{i+1} = x_i + step * \delta$$

It is in the mechanisms to choose the size of *step* and to decide when to terminate the iteration that Gauss-Newton methods differ. Indeed, though I have tried several times, I find the very convoluted code behind `nls()` very difficult to decipher. Unfortunately, its authors have now (as far as I am aware) all ceased to maintain the code.

We could implement the methods using the equations above. However, the accumulation of inner products in $J^T J$ occasions some numerical error, and it is generally both safer and more efficient to use matrix decompositions. In particular, if we form the QR decomposition of $J$

$$QR = J$$

where Q is an orthonormal matrix and R is Right or Upper triangular, we can easily solve

$$R\delta = -Q^T r$$

for which the solution is also the solution of the Gauss-Newton equations. But how do we get the Marquardt stabilization?

If we augment $J$ with a square matrix $\sqrt{\lambda D}$ whose diagonal elements are the square roots of $\lambda$ times the diagonal elements of $D$, and augment the vector $r$ with $n$ zeros where $n$ is the column dimension of $J$ and $D$, we achieve our goal.

Typically we can use $D = I_n$ (the identity of order $n$), but Marquardt (1963) showed that using the diagonal elements of $J^T J$ for $D$ results in a useful implicit scaling of the parameters. Nash (1977) pointed out that on computers with limited arithmetic (which now are rare since the IEEE 754 standard appeared in 1985), underflow might cause a problem of very small elements in $D$ and proposed adding $\phi I_n$ to the diagonals of $J^T J$ before multiplying by $\lambda$ in the Marquardt stabilization. This avoids some awkward cases with very little extra overhead. It is accomplished with the QR approach by appending $\sqrt{\phi\lambda}$ times a unit matrix $I_n$ to the matrix already augmented matrix. We must also append a further $n$ zeros to the augmented $r$.

[2021-5-27] JN added some notes on LU and Cholesky to "Variety" document.

**======== End of extract on approaches to solving Gauss Newton equations =========**

## Early explorations and setup of project [2021-5-24]

This is essentially a diary entry. (?? AB you may want to add some notes too.)

- AB working to secure his computer, set up VirtualBox and prepare a linux guest VM
- AB has been looking at documents JN posted. Gitlab repo seems to be working OK. On comment from HT, JN discovered (very overdue!) that Rstudio support for git is very useful.
- JN set up nls-flowchart.txt for use by team. This is looking at R-devel.tar.gz downloaded May 18, 2021. It is likely we can use any recent version, though at some point we may need to be more precise. The purpose of this is to describe what is going on in the nls() operations.
- JN added document **VarietyInNonlinearLeastSquaresCodes.Rmd**. This is tangential to the project, but is an attempt to provide a panorama of the different ways in which nonlinear least squares software has been set up over the years. JN will respond to questions from other team members to try to make this explanatory document helpful in understanding the existing nls() code and proposed changes.
- Section added above "From `nlsr` vignette `nlsr-devdoc.Rmd`"

- In the "VarietyInNonlinearLeastSquaresCodes.Rmd" document, JN added a section on different approaches to solving LINEAR least squares problems, since the inner iteration of almost all nonlinear least squares methods involves such a sub-problem. AB has been looking at this. It is currently a messy DRAFT section at 2021-5-25.
- Meet scheduled for May 25;

## Attempts to understand numericDeriv() [2021-5-24 to 27]

JN has found it quite difficult to get a good understanding of the `numericDeriv()` function that appears in `nls.R` and calls material in `nls.c`. That this has proved difficult to use suggests the documentation is less than wonderful. However, the main reason is to set up tests and timings and compare to other approaches to computing the Jacobian in nonlinear least squares. ??Gut feeling: numericDeriv is really only set up to compute a Jacobian. ?? Note: there is a "central" option. Can this be used with nls()? Why or why not?

A file ExDerivs.R has been added to the repo.

[2021-5-26] JN sorted out one approach that allows numricDeriv, nlsr::model2rjfun and numDeriv::jacobian to be compared. [2021-5-27] DerivsNLS.Rmd first draft done. Problems with system() and system2() and calling bash from rmarkdown document led to email sent to Yihui Xie, as the output of a Linux command (inxi -F) to get machine properties was returned with extra characters.

## Tangential issues raised during the project

### Output from `bash` scripts in rmarkdown files

In trying to capture the output of the Linux command-line program `inxi` (which reports hardware and software information), JN found rmarkdown failed to knit correctly.

```
system('inxi -F >t.txt')
```

and

```
system2("inxi -F >tlinux.txt")
```

both failed, as did specifying the chunk as `{bash inxi2}` and trying `inxi -F >t2.txt`.

```
The outputs had strange characters that were indecipherable by the LaTEX processor
and caused a crash. Indeed files of substantially differing sizes were generated.

We also tried
```

system2("inxi," "-F," stdout="tlinux2.txt")

```
The issue was reported on community.rstudio.

\url{https://community.rstudio.com/t/bash-output-causes-trouble-for-latex-and-hence-no-pdf/106502}

where a workaround is reported. This puts the inxi command in a script which uses the preamble
`/#!/bin/sh`. That is, we use the `sh` shell. For example, let us call it myscript.sh and it
contains
```

#!/bin/sh # myscript.sh – push hw info to thw.txt one directory above inxi -F >../thw.txt echo "done"

```
This command script is then run from R using
```

system("mate-terminal -e ./myscript.sh") "'

JN tested this with mate-terminal, xterm and konsole as the terminal programs.

## References


Hartley, H. O. 1961. "The Modified Gauss-Newton Method for Fitting of Nonlinear Regression Functions by Least Squares." *Technometrics* 3: 269–80.

Levenberg, Kenneth. 1944. "A Method for the Solution of Certain Non-Linear Problems in Least Squares." *Quarterly of Applied Mathematics* 2: 164--168.

Marquardt, Donald W. 1963. "An Algorithm for Least-Squares Estimation of Nonlinear Parameters." *SIAM Journal on Applied Mathematics* 11 (2): 431–41.

Nash, John C. 1977. "Minimizing a Nonlinear Sum of Squares Function on a Small Computer." *Journal of the Institute for Mathematics and Its Applications* 19: 231–37.

———. 1979. *Compact Numerical Methods for Computers : Linear Algebra and Function Minimisation*. Book. Hilger: Bristol.