MATLAB Code for fitting "Envelope Models for Parsimonious and Efficient Multivariate Linear Regression"

R. Dennis Cook*

July 29, 2009

Fitting of the envelope model is included as part of a larger MATLAB package **LDR** for likelihood-based sufficient dimension reduction that is being developed by R. D. Cook, Liliana Forzani and Diego Tomassi. Envelope fitting is included in the package because its objective function is similar to ones encountered in sufficient dimension reduction. This document covers only the fitting of multivariate envelope models; computing for sufficient dimension reduction are discussed elsewhere. Familiarity with MATLAB is assumed.

Start by unzipping the zip file. The directory created includes a number of MATLAB files and subdirectories. The only directory relevant to this application is titled 'mlm' for multivariate linear model.

Start MATLAB in the unzipped directory and type the command 'setpaths'. Then create an $n \times r$ matrix Y of responses and an $n \times p$ matrix of predictors. The next

*School of Statistics, University of Minnesota, 313 Ford Hall, 224 Church Street SE, Minneapolis, MN 55455. email: dennis@stat.umn.edu. Research for this work was supported in part by National Science Foundation Grant DMS-0704098. The paper "Envelope Models for Parsimonious and Efficient Multivariate Linear Regression" by R. D. Cook, B. Li and F. Chiaromonte is to appear in *Statistica Sinica* with discussion. A preprint is available at http://www.stat.umn.edu/ dennis/RecentArticles/CLC.pdf.

step is construct an estimate G of $\mathbf{\Gamma}$. After that, all subsequent commands will use Y, X, and G.

The basic command for finding G is `[GX,G,L,dhat] = mlm_fit(Y,X,dim)`, where `dim` stands for the dimension. The output consists of

- The estimate `G` of $\mathbf{\Gamma}$.
- The value `L` of the log likelihood at the MLE's
- The estimated dimension `dhat` of the envelope.
- `GX` is an exploratory quantity and not used routinely.

The following options are avaiable for `dim`:

- If `dim` is set to an interger d, eg., `mlm_fit(Y,X,2)`, then `dhat` $= d$ and the envelope model is fitted with the specified dimension.
- If `dim` is set to `'aic'` or `'bic'` (quotes included) then the indicated information criterion AIC or BIC is used to estmimate d. These commands may take a long time, depending on the size of the regression.
- If `dim` is set to the pair `'lrt', .05`, eg,. `mlm_fit(Y,X,'lrt', .05)`, then likelihood ratio testing at level .05 is used to estimate d.

The following commands are available following the initial fit that produces `G`.

- `[beta,S,Sfit,Sres] = mlm_fmpars(Y,X)` returns the parameter estimates from the full multivariate linear model ($d = \min(r, p)$). `beta` is the estimated coefficient matrix, `S` is the marginal covariance matrix of the responses, `Sfit` is the covariance matrix of the fitted vectors, and `Sres` is the covariance matrix of the residual vectors.
- `mlm_fmses(Y,X)` returns the matrix of standard errors of the elements of `beta` from the fit of the full multivariate linear model.

- `[betaem,eta,Omega,Omega0,S1,S2] = mlm_empars(Y,X,G)` returns the estimated parameters from the fit of the envelope model. `betaem = G x eta` is the estimated coefficient matrix `Omega` and `Omega0` are as described in the paper `S1` is the estimate of $\boldsymbol{\Sigma}_1$ and `S2` is the estimate of $\boldsymbol{\Sigma}_2$.

- `mlm_emses(Y,X,G)` returns the matrix of standard errors of the elements of `betaem` from the fit of the envelope model.

- `mlm_seratios(Y,X,G)` returns the matrix of ratios of standard errors of the full model and envelope model estimates of $\boldsymbol{\beta}$. The ratios are (full model se's)/(envelope model se's).

The following output reproduces the analysis of the wheat protein data. These data are available in the `data` directory, so the commands following the `>>` prompt should be usable immediately after the `setpaths` command.

```
>> data=load('./data/wheatprotein.txt');
>> Y = data(:,1:6);
>> X = data(:,8);
>> [GX,G,L,dhat]=mlm_fit(Y,X,'lrt',.05);
      MODEL SELECTION DONE...
>> G
G =
   -0.1243
    0.5223
    0.4301
   -0.6979
    0.0702
   -0.1866
>> L
L =
```

```
 -395.8468

>> dhat

dhat =

     1

>> mlm_fmses(Y,X)

ans =

   69.0626
   56.8296
   61.0742
   67.0707
   96.4705
   38.0740

>> mlm_emses(Y,X,G)

ans =

    2.4631
    3.0888
    2.5888
    4.1166
    1.4661
    5.8980

>> mlm_seratios(Y,X,G)

ans =

   28.0389
   18.3983
   23.5916
   16.2928
   65.7999
    6.4555
```

```
>> [betaem,eta,Omega,Omega0,S1,S2] = mlm_empars(Y,X,G);
>> betaem
betaem =
   -1.0644
    4.4730
    3.6839
   -5.9770
    0.6013
   -1.5986
>> Omega
Omega =
    7.8762
>> S1
S1 =
    0.1217   -0.5112   -0.4210    0.6831   -0.0687    0.1827
   -0.5112    2.1483    1.7693   -2.8706    0.2888   -0.7678
   -0.4210    1.7693    1.4572   -2.3642    0.2379   -0.6323
    0.6831   -2.8706   -2.3642    3.8359   -0.3859    1.0259
   -0.0687    0.2888    0.2379   -0.3859    0.0388   -0.1032
    0.1827   -0.7678   -0.6323    1.0259   -0.1032    0.2744
>> S2
S2 =
   1.0e+03 *
    1.1951    0.9802    1.0551    1.1561    1.5413    0.6358
    0.9802    0.8070    0.8673    0.9497    1.2657    0.5293
    1.0551    0.8673    0.9333    1.0225    1.3688    0.5668
    1.1561    0.9497    1.0225    1.1229    1.5262    0.6195
    1.5413    1.2657    1.3688    1.5262    2.3246    0.8375
```

```
    0.6358    0.5293    0.5668    0.6195    0.8375    0.3628

>> eig(S1)

ans =

   -0.0000

    7.8762

   -0.0000

   -0.0000

   -0.0000

    0.0000

>> eig(S2)

ans =

   1.0e+03 *

    6.5166

    0.2083

    0.0201

    0.0000

    0.0004

    0.0003
```