Screening Paper: Tunning Parameters Summary

Dantzig Selector

1. δ

   - Occurs in
   $$\min_{\hat{\boldsymbol{\beta}} \in R^k} ||\hat{\boldsymbol{\beta}}|| \text{ subject to } ||\boldsymbol{X}'r||_{l_\infty} \leq \delta$$
   where $r = y - \boldsymbol{X}\hat{\boldsymbol{\beta}}$, $||\boldsymbol{a}|| = \sum |a_i|$, and $||\boldsymbol{a}||_{l_\infty} = \max |a_i|$. (pg 3)
   - Paper suggestion: estimate by running DS over a grid for δ's (no indication for size or coarseness) and pick the best model and δ via chosen criterion (e.g. mAIC $= n\ log(RSS/n) + 2p^2$).
   - Currently implemented grid: 0 to 15 by 0.5.
   - June 11^{th}, 2009 – Modified implemented procedure to tune δ: See the comment for γ below.

2. γ

   - Occurs in "...estimate $I = \{i : \beta_i \neq 0\}$ with $\hat{I} = \{i : \hat{\beta}_i > \gamma\}$ for some $\gamma \geq 0$...". (pg 4)
   - Paper suggestion: let γ be equal to 1/10 of the largest $|\hat{\beta}_i|$ in the model with $\delta = 0$. (pg 5)
   - Currently implemented value: used paper's suggestion.
   - June 11^{th}, 2009 – Modified implemented procedure to tune γ: Let $\gamma_{min} = 3$ to $\gamma_{max} = 20$ by $\gamma_{step} = 1$ and 0 to $\delta_{max} = max(|X'y|)$ by $\delta_{step} = \delta_{max}/200$ be grid values for γ and δ, respectively. For each γ value an optimal δ (i.e the best model) via mAIC will be chosen. Then, the models corresponding to each γ will be further distinguished via mAIC criterion in order for the final model to be chosen.
   - June 27^{th}, 2009 – γ_{max} cannot be arbitrarily large. When γ_{max} is too large, the DS algorithm returns all NULL values. By setting γ_{max} too large, we forcibly make all effects inactive and the algorithm has nothing to do. Thus, the algorithm starts at $\gamma_{min} = 3$ and increases γ by 1 until either first NULL value is returned or $\gamma_{max} = 20$ is reached. Then the mAIC criterion is computed for the models corresponding to the valid γ values.
   - July 1^{st}, 2009 – Because of the same reason as in June 27^{th}, 2009 note, $\gamma_{min} = 1$. γ_{min} cannot be so large that algorithm will not find anything for the smallest value of γ on the grid.
   - July 11^{th}, 2009 – Let γ_{min} is reset to 3. If the procedure returns a NULL value for $\gamma = 3$ then the result is recorded as if procedure did not find any active effects. δ grid is 0 to $\delta_{max} = max(|X'y|)$ by $\delta_{step} = \delta_{max}/100$.

SCAD

1. a

   - Occurs in "... smoothly clipped absolute deviation (SCAD), whose first order derivative is defined by

     $$\frac{\partial \phi}{\partial \beta} = \lambda\{I(\beta \leq \lambda) + \frac{a\lambda - \beta}{(a-1)\lambda} I(\beta > \lambda)\}$$

     for some $a > 2$ and $\beta > 0$ with $p_\lambda(0) = 0$".
   - Paper suggestion: $a = 3.7$.
   - Currently implemented value: $a = 3.7$.

2. λ

   - Occurs in – see the above entry for a.
   - Paper suggestion: obtain λ by minimizing corresponding GCV scores, i.e. minimize

     $$GCV(\lambda) = \frac{(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})}{n(1 - \ \mathrm{tr}(H(\lambda))/n)^2}$$

     where $\mathrm{tr}(H(\lambda)) = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X} + n\Sigma_\lambda(\hat{\boldsymbol{\beta}}^0))^{-1}\boldsymbol{X}'$,
     $\Sigma_\lambda(\hat{\boldsymbol{\beta}}^0)) = \mathrm{diag}(\phi'(|\hat{\boldsymbol{\beta}}_1^0|)/|\hat{\boldsymbol{\beta}}_1^0|, \ldots, \phi'(|\hat{\boldsymbol{\beta}}_p^0|)/|\hat{\boldsymbol{\beta}}_p^0|)$.
   - Currently implemented value: used paper's suggestion.

3. α-in and α-out

   - Occur in Stepwise Selection which is performed before SCAD analysis in order to reduce the number of explanatory variables so that the initial value for $\boldsymbol{\beta}$ can be obtained vie ls estimate.
   - Paper suggestion: set α-in and α-out to the large value of so that all active factors are included into selected model. Set α-in and α-out to 0.1. (pg 139 Li and Lin (2002))
   - Currently implemented values: 0.1 for both.

4. Size of the subset of the variables chosen by Stepwise Selection which will be used in the estimation of the error variance

   - Occurs in – see Figure 1 below; The line of interest is x1=x0(:,1:7); Why 7?
   - Paper suggestion: Not sure that I've seen one. However, Professor Li's Matlab code came (via Qing Liu) with a little manual which says the following:
     - "Degrees of freedom are needed to estimate error variance and, therefore, the number of columns in the model matrix cannot exceed the number of runs. As shown in *super.m* program in Professor Li's codes (see Figure 1), a subset of variables (columns) (less than number of runs) in $\boldsymbol{X}0$ is taken from $\boldsymbol{X}1$

which is then used to obtain the error variance estimate s0. Different choices of $\boldsymbol{X}1$ lead to different realizations of s0, which give rise to different models fitted by SCAD procedure. For bigger models include more variables in $\boldsymbol{X}1$, and vice versa."

- Currently implemented: max(# variables chosen by Stepwise - 2, 1)
- June 11^{th}, 2009 – Modified implemented value: For a given number of factors, f, P(main effect active) = q_{me}, and P(interaction effect active) = q_{int}, the expected number of active factors is $fq_{me} + \binom{f}{2}q_{int}$ (see Table 1). The implemented value is $\min((fq_{me} + \binom{f}{2}q_{int} + 1)^{\uparrow}$, # variables chosen by Stepwise) where $^{\uparrow}$ indicates that the number has been rounded up to the nearest integer.

Table 1: Expected Number of Active Factors for f and q_{me}

	10	15	20	25	30
0.05	1.27	2.54	4.23	6.35	8.90
0.10	2.31	4.55	7.52	11.22	15.64
0.15	3.35	6.57	10.81	16.08	22.38
0.20	4.39	8.59	14.11	20.96	29.14

Group Screening

1. $R^2_{increment}$

   - Occurs in R^2 selection procedure in both stages of Group Screening (unless some other procedure is used to identify active effects). Given the effects already in the model, another effect is included if, upon it's inclusion, R^2 rises more than $R^2_{increment}$ percent.
   - Currently implemented value: 15%. This value has been chosen after doing a simulation to examine the sizes of increments in R^2 under different scenarios considered for the screening paper.
   - Note: We were aware of the significance of the size of $R^2_{increment}$ on the overall performance of Group Screening. Upon brief examination of the preliminary screening results for $f = 25$ and $q_{me} = 0.10$ case, it can be seen that the current value of $R^2_{increment}$ severely hinders Group Screening in its ability to perform well. This value should probably be adjusted for different f's by assuming effect sparsity such that it adjusts for the different f's and the number of factors we expect to be active.
   - June 11^{th}, 2009 – Modified implemented value: $99/(fq_{me} + \binom{f}{2}q_{int} + 1)^{\uparrow}$.

2. $R^2_{threshold}$

Figure 1: Part of *super.m* Matlab code from Professor Li

```
  load 'superas.dat';
x=superas(:,1:23);
x=[ones(size(x,1),1),x];
y=superas(:,24);

fin=0.1;
fout=0.1;

J0=stepwise(x,y,fin,fout)

x0=x(:,J0);

x1=x0(:,1:7);
beta0 = inv(x1'*x1)*x1'*y;

s0=sum((y-x1*beta0).^2)/(14-length(beta0));

lambda1 = gcv_scad([x0,y],s0)

betahat = inv(x0'*x0)*x0'*y;
%std0=diag(inv(x0'*x0))*sqrt(s0)
[beta,std] = scad(betahat,lambda1,[x0,y],s0)
```

- Occurs in R^2 selection procedure in both stages of Group Screening (unless some other procedure is used to identify active effects). The selection stops once R^2 for the current model is over $R^2_{threshold}$ value or when the increment is below $R^2_{increment}$ value.
- Currently implemented value: 90%.
- June 11^{th}, 2009 – Modified implemented value: 100% (essentially getting rid off this tuning parameter).

3. k

- Occurs in k-exchange algorithm and it represents the number of least critical points in a design which will be exchanged in an attempt to improve the design under a chosen criterion. For example, if the desired design regards D-optimality, the points which will be exchanged are the ones which have the smallest values of the *deletion function* $d_D(\boldsymbol{x}_i) = f'(\boldsymbol{x}_i)(\boldsymbol{X}'\boldsymbol{X})^{-1}f(\boldsymbol{x}_i)$ (Meyer and Nachtsheim 1995 pg 61).
- Paper suggestion: Cook and Nachtsheim (1992) via Meyer and Nachtsheim (1995) pg 61 suggest $k \leq n/4$. Dave suggested in one of his emails setting $k = n$ which

would result in modified Fedorov algorithm (Cook and Nachtsheim 1980).

- Currently implemented: $k = n/4$.
- July 1^{st}, 2009 – $k = n$.

4. Grouping Scheme - we've settled here on 5 groups of equal size.

Forward

1. α-in

- This is the usual value in the Forward selection procedure
- Currently implemented values: α-in = 0.05.
- June 11^{th}, 2009 – Run Forward Selection for each $\alpha\text{-in} \in \{0.05, 0.10, 0.15\}$ resulting in 3 competing models and choose the final model via mAIC.

Stepwise$_{CV}$

1. numPts

- numPts represents the number of points which are dropped one at a time during cross-validation.
- Current implementation: $0.1n$ rounded to the nearest integer.

2. α-in and α-out

- These are the usual values in the Stepwise procedure.
- The Stepwise$_{CV}$ procedure proceeds by creating a grid of values for α-in and α-out such that $\alpha\text{-in} \geq \alpha\text{-out}$. Then the numPts CV points are randomly chosen. Then for each (α-in, α-out) pair, each of the numPts are dropped, Stepwise is run, and the prediction is made for the dropped point based on the factors identified by the procedure as active. Finally, MSPE based on the numPts components is computed for each (α-in, α-out) pair. The pair which minimizes MSPE is chosen as the winner and the Stepwise is run one more time on the full data set to obtain the final set of factors deemed active.
- Current implementation: $(\alpha\text{-in}, \alpha\text{-out}) \in \{(0.05, 0.05), (0.10, 0.05), (0.10, 0.10), (0.15, 0.05), (0.15, 0.10), (0.15, 0.15)\}$
- June 11^{th}, 2009 – Modified implemented value: Eliminate cross-validation (and numPts). Run Stepwise for each $(\alpha\text{-in}, \alpha\text{-out}) \in \{(0.05, 0.05), (0.10, 0.05), (0.10, 0.10), (0.15, 0.05), (0.15, 0.10), (0.15, 0.15)\}$ resulting in 6 competing models and choose the final model via mAIC.

LARS and **LASSO**

1. These two procedures are implemented using R package *lars* which implements LARS algorithm as described in http://www-stat.stanford.edu/~hastie/Papers/LARS/LeastAngle_2002.pdf. The algorithm requires no user provided tuning parameters.

2. When $f >> n$ LARS algorithm terminates when $n - 1$ variables have been declared active. At the kth step, at most $n - 1$ factors are regarded active. Thus, each step produces a competing model. After the LARS algorithm is done the final model is chosen via mAIC.