Statistical Methods III: Spring 2013

Jonathan Wand

Stanford University

QP + constrained inference

Outline

1. Quadratic programming
2. Testing inequality constraints
3. Ordered means

OLS

Scalar notation

$$\arg\min_{\beta} \sum (y_i - x^\top \beta)^2$$

Matrix notation

$$\arg\min_{\beta} (Y - X\beta)^\top (Y - X\beta)$$

Matrix notation expansion

$$\arg\min_{\beta} \beta^\top X^\top X\beta - 2Y^\top X\beta + Y^\top Y$$

- No constraints on β

Reviewing (and rewriting) OLS

Matrix notation expansion

$$\arg\min_{\beta} \beta^\top X^\top X\beta - 2Y^\top X\beta + Y^\top Y$$

multiply by 1/2, and drop YY (why allowed?)

$$\arg\min_{\beta} \frac{1}{2}\beta^\top X^\top X\beta - Y^\top X\beta$$

If we write in form of,

$$\arg\min_{\beta} d^\top \beta + \frac{1}{2}\beta W\beta$$

then

- $W = X^\top X$. What is W^{-1}?
- $d = -Y^\top X$

OLS - with restrictions

OLS solution places no restriction on β.
The optimization problem

$$\arg\min_{\beta} \frac{1}{2}\beta^\top X^\top X\beta - Y^\top X\beta$$

restricted inequality constraints on β, for example,

$$\beta \geq 0$$

can be solved with "quadratic programing"

Quadratic programming

Definition (Quadratic programming problem)

Let $\beta \in R^n$, W be symmetric $n \times n$ matrix,

$$\arg\min_{\beta} d^\top \beta + \frac{1}{2}\beta^\top W \beta$$

subject to

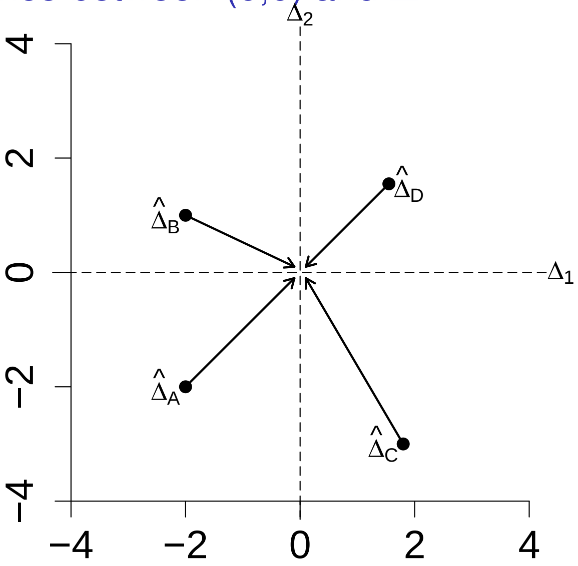
$$R_1^\top \beta \geq b_1$$

$$R_2^\top \beta = b_2$$

Notes

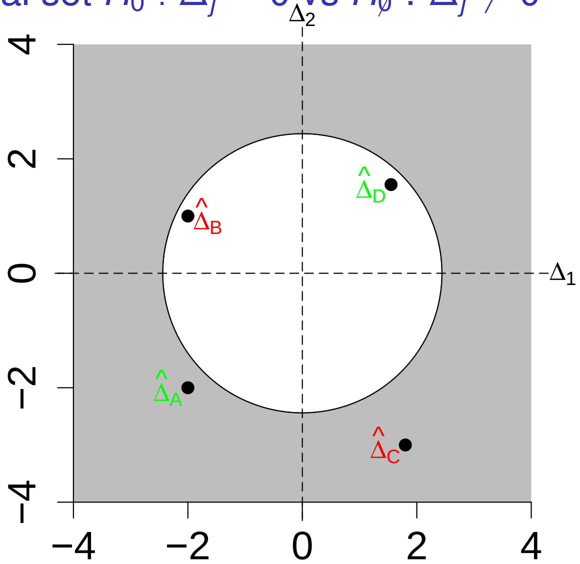
- d, W, R, and b are fixed for a given optimization
- β is unknown

Distance between (0,0) and $\hat{\Delta}$



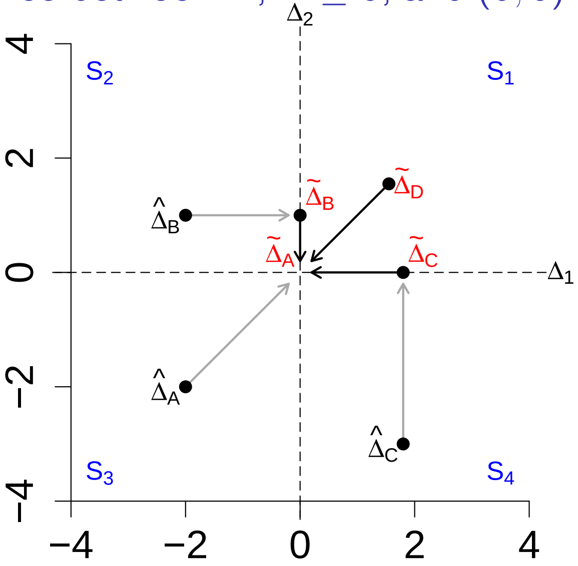
$\Delta_1^2 + \Delta_2^2 \sim \chi_2^2$

Critical set $H_0 : \Delta_j = 0$ vs $H_{\emptyset} : \Delta_j \neq 0$



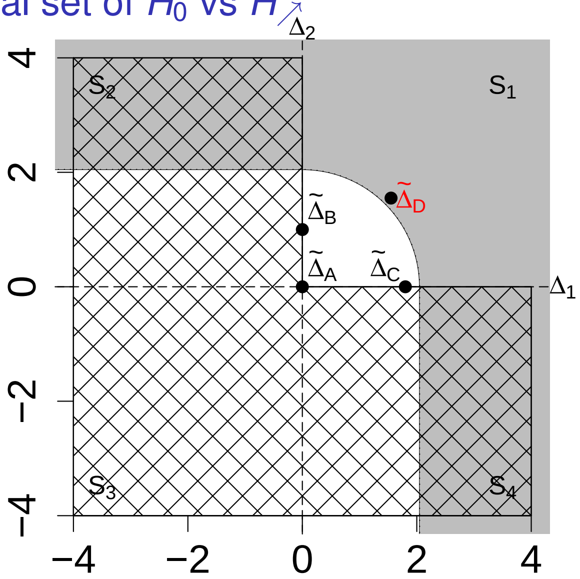
Critical value $\chi^2_{2,\alpha=.05} = 6$, so $\sqrt{\Delta_1^2 + \Delta_2^2} = \sqrt{6} =$ radius

Distance between $\hat{\Delta}$, $\tilde{\Delta} \geq 0$, and $(0, 0)$



$\tilde{\Delta}_A \sim 0 \quad \tilde{\Delta}_B, \tilde{\Delta}_C \sim \chi^2_1 \quad \tilde{\Delta}_D \sim \chi^2_2$

Critical set of H_0 vs $H_{\nearrow}$



$$P(\bar{\chi}^2_{+,\alpha} > \bar{c}) = 1/4 P(\chi^2_2 > \bar{c}) + 1/2 P(\chi^2_1 > \bar{c}) + 1/4 = \alpha$$

Constrained estimation: dimensions and distance

Constrained estimates $\tilde{\Delta}$, subject to $\Delta > 0$

$$\tilde{\Delta} = \begin{cases} (\hat{\Delta}_1, \hat{\Delta}_2) & \text{if } \hat{\Delta}_1 \geq 0 \text{ and } \hat{\Delta}_2 \geq 0 \\ (\hat{\Delta}_1, 0) & \text{if } \hat{\Delta}_1 \geq 0 \text{ and } \hat{\Delta}_2 < 0 \\ (0, \hat{\Delta}_2) & \text{if } \hat{\Delta}_1 < 0 \text{ and } \hat{\Delta}_2 \geq 0 \\ (0, 0) & \text{if } \hat{\Delta}_1 < 0 \text{ and } \hat{\Delta}_2 < 0 \end{cases}$$

How many free parameters? 2 (S1), 1 (S2, S4), or 0 (S3).
What is probability of being "far" from H_0? By quadrant:

$$\begin{aligned} P(\tilde{\Delta}_1^2 + \tilde{\Delta}_2^2 < c' \mid \hat{\Delta} \in S_1) &= P(\chi_2^2 < c') \\ P(\tilde{\Delta}_1^2 + 0 < c' \mid \hat{\Delta} \in S_2) &= P(\chi_1^2 < c') \\ P(0 + \tilde{\Delta}_2^2 < c' \mid \hat{\Delta} \in S_4) &= P(\chi_1^2 < c') \\ P(0 + 0 < c' \mid \hat{\Delta} \in S_3) &= 1 \end{aligned}$$

Constrained estimation: dimensions and distance

In this simple example, each of quadrants is equally likely under H_0

$$
\begin{aligned}
P(\tilde{\Delta}_1^2 + \tilde{\Delta}_2^2 < c' \mid \hat{\Delta} \in S_1) &= P(\chi_2^2 < c') \\
P(\tilde{\Delta}_1^2 + 0 < c' \mid \hat{\Delta} \in S_2) &= P(\chi_1^2 < c') \\
P(0 + \tilde{\Delta}_2^2 < c' \mid \hat{\Delta} \in S_4) &= P(\chi_1^2 < c') \\
P(0 + 0 < c' \mid \hat{\Delta} \in S_3) &= 1
\end{aligned}
$$

So for given α, solve for c'

$$
P(\bar{\chi}^2 < c') = 1/4P(\chi_2^2 < c') + 1/2P(\chi_1^2 < c') + 1/4 = 1 - \alpha
$$

Distribution of hypothesis tests

- Equality restriction versus unconstrained
  - Fixed difference in dimensionality of models

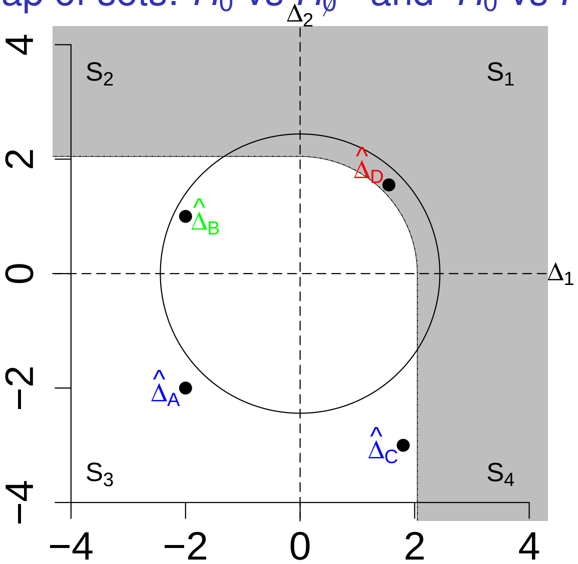
$$Pr(-2[L(\tilde{\theta}) - L(\hat{\theta})] > c) = Pr(\chi^2_r > c)$$

- But number of free $-$ restricted parms is stochastic if
  - Equality restrictions versus inequality restriction
  - Inequality restricted vs unconstrained
  - Inequality restricted versus additional inequalities

$$\begin{aligned} Pr(-2[L(\tilde{\theta}) - L(\hat{\theta})] > c') &= Pr(\bar{\chi}^2 > c') \\ &= \sum_{k=1}^{K} w_k Pr(\chi^2_k > c') \end{aligned}$$

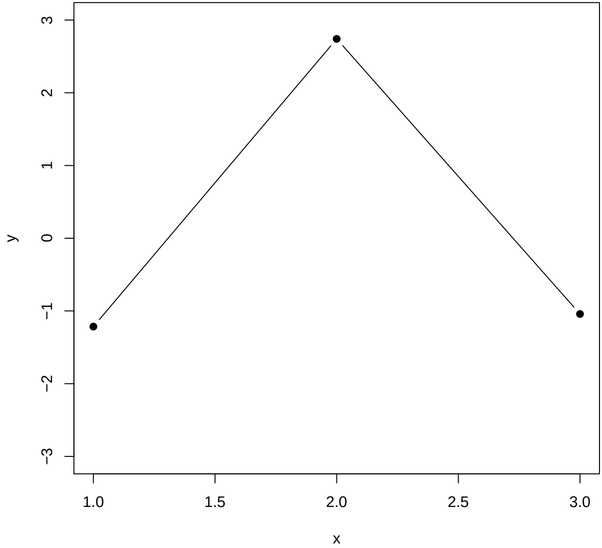
where w_k is the probability of having a difference of k degrees of freedom between models

Overlap of sets: H_0 vs $H_\emptyset$ and H_0 vs $H_\nearrow$

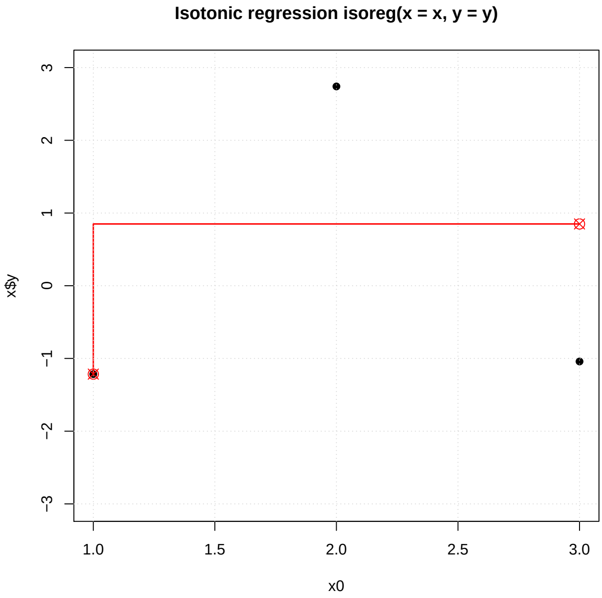


$\sqrt{\chi^2_{2,\alpha}} = radius_\emptyset = 2.45$ vs $\sqrt{\bar{\chi}^2_{+,\alpha}} = radius_\nearrow = 2.05$

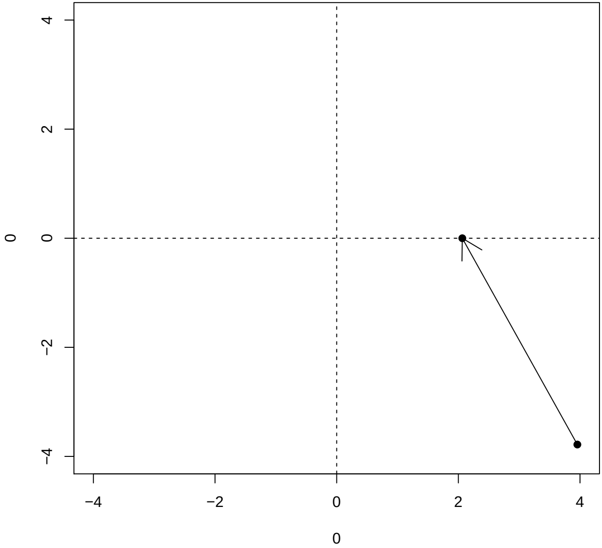
Sequence of means



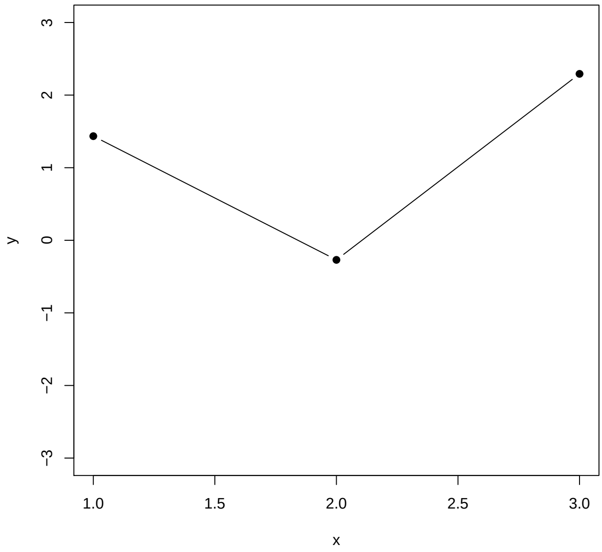
Best fitting mono inc.



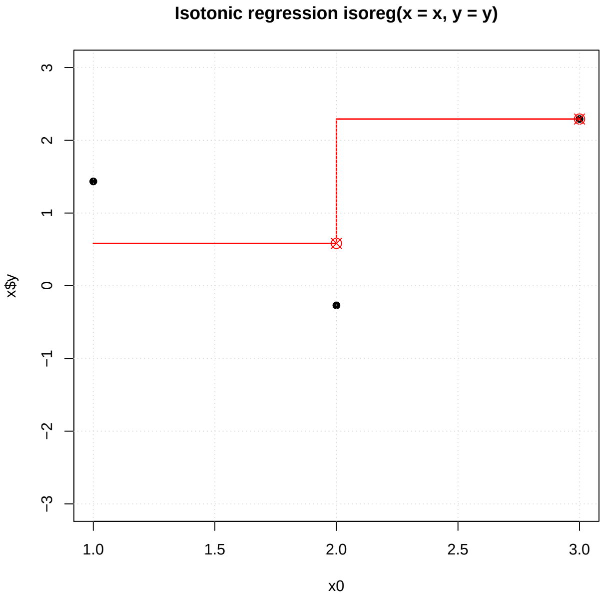
Diff in adjacent mean, unconstrained to constrained



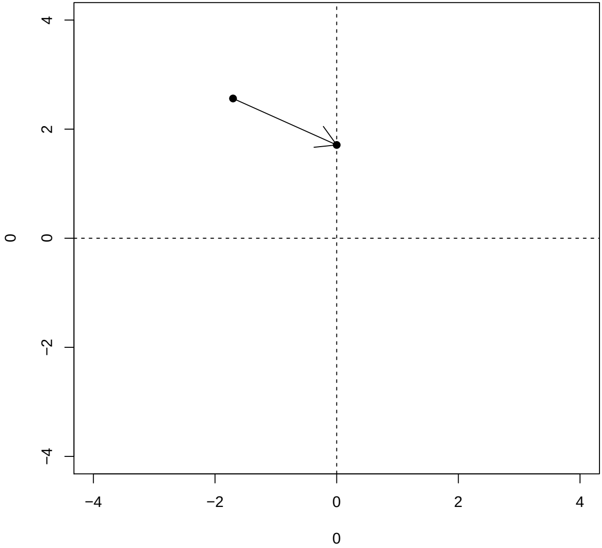
Sequence of means



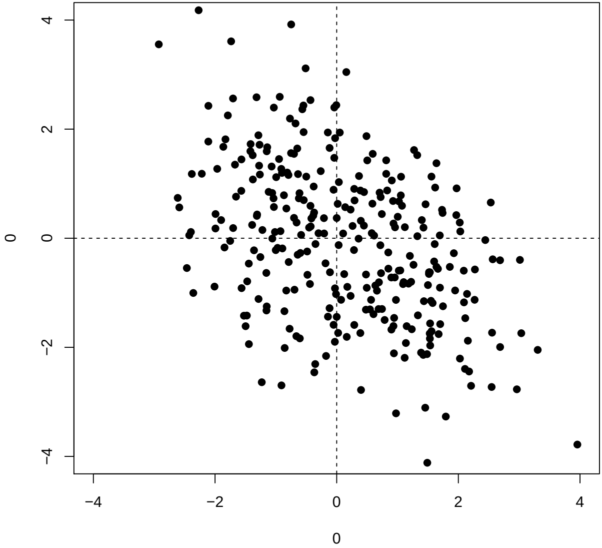
Best fitting mono inc.



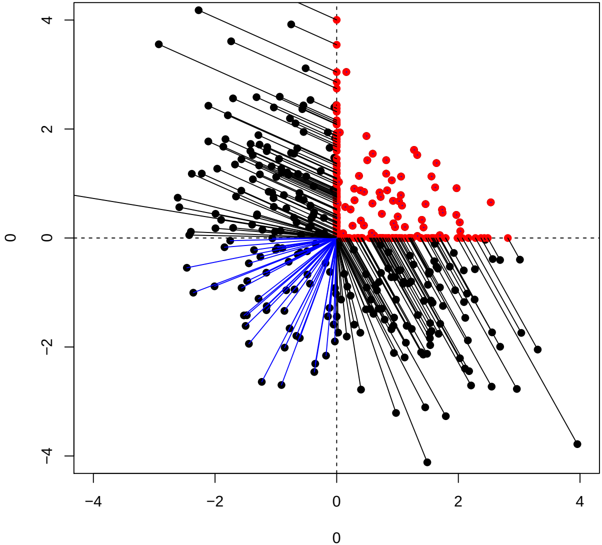
Diff in adjacent mean, unconstrained to constrained



Lots of draws of unconstrained diff means



Constrained v mono inc diff

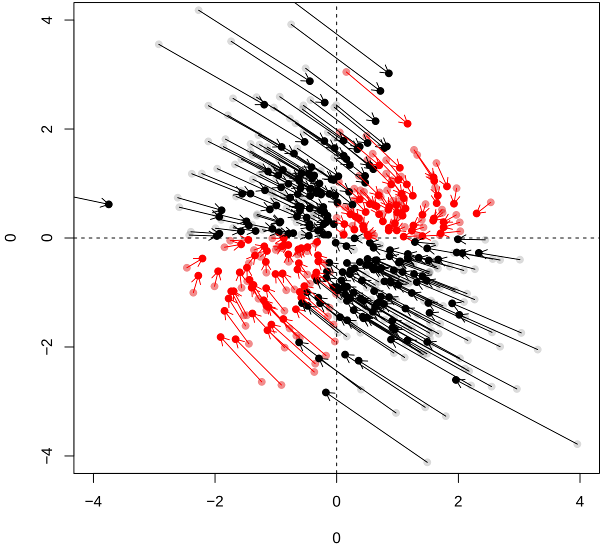


Constrained v mono inc diff

In simulation,

- could you determine w, probability of model dimensionality
- what is the distribution of the fits in constrained?

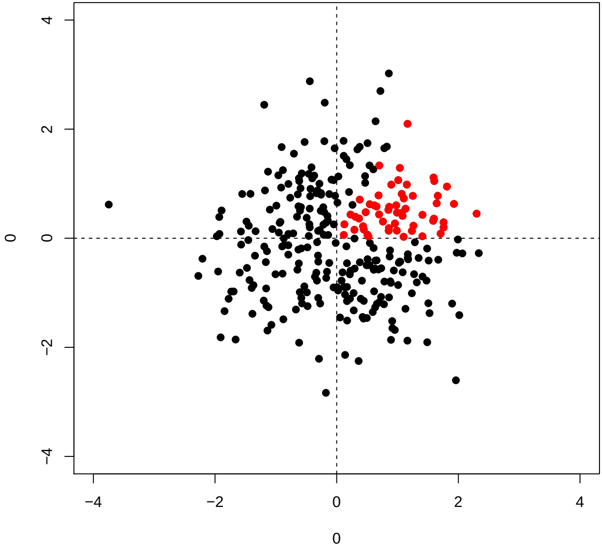
Transformation via cholesky



Transformation via cholesky

```
Y <- matrix(rnorm(3*n),ncol=3)
d <- t(apply(Y,1,diff))
z <- t(apply(Y,1,function(x) isoreg(x)$yf ))
d2<- t(apply(z,1,diff))

(V <- cov(d))
Vi<- solve(V)
A <- t(chol(Vi))
Ai<- solve(A)
d3 <- d %*% A
```

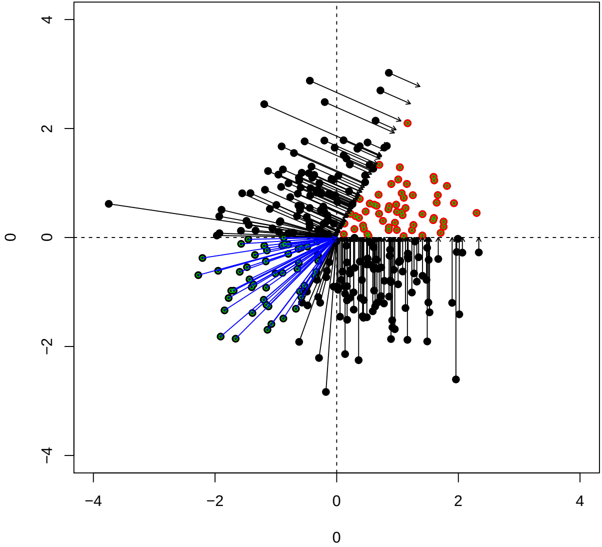
... now iid bivariate normal



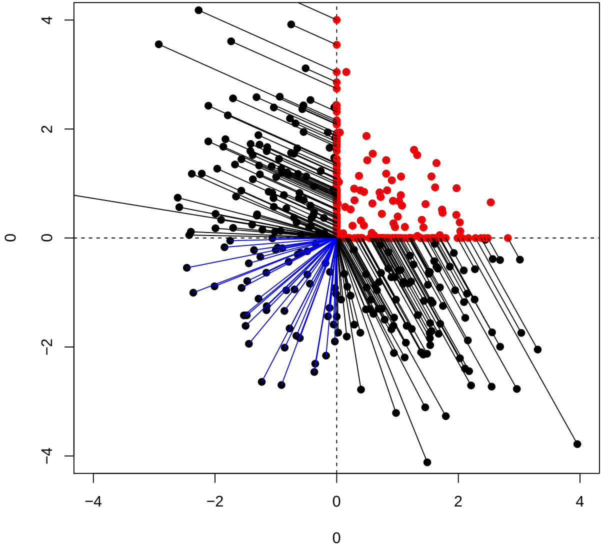
Transformed optimization problem...

```
d <- t(apply(Y,1,diff))
d2<- t(apply(z,1,diff))
(V <- cov(d))
Vi<- solve(V)
A <- t(chol(Vi))
Ai<- solve(A)
d3 <- d %*% A

qp2 <- function(X)
  solve.QP( Dmat = Ai%*%Vi%*%t(Ai),
            dvec = X %*% A ,
            Amat =  t(R %*% t(Ai)) ,
            bvec = c(0,0))

X   <- d[i,]
d4 <- qp2(X)
```

Constrained fit on transformed means



Constrained v mono inc diff



Transformed vs untransformed

In simulation,

- R and V determine cone/constraint
- note: perpendicular: closest fitting point
  compare with untransformed
- could you determine *w*, probability of model dimensionality (any different from untransformed?)
- what is the distribution of the fits in constrained?