Optimal solutions to geodetic inverse problems in statistical and numerical aspects

Jianqing Cai
Geodätisches Institut, Universität Stuttgart

Session 6: Theoretische Geodäsie
Geodätische Woche 2010
05.–07. Oktober 2010, Messe Köln

Main Topics

1. Geodetic data analysis
2. Fixed effects (G-M, Ridge, BLE, α- weighted homBLE)
3. Regularizations
4. Mixed model (Combination method)
5. Conclusions and further studies

1. Geodetic data analysis

- Many diffenent developed independently in separate disciplines.
- Some clear differences in terminology, philosophy and numerical implementation remain,
- due to tradition and lack of interdisciplinary communication.

- Statistical aspect: Estimation and inference
  - the data are modeled as stochastic
  - Standard statistical concepts, questions, and considerations such as bias, variance, mean-square error, identifiability, consistency, efficiency and various forms of optimality can be applied
  - ill-posed problems is related to the rank defect, etc.
  - Biased and unbiased estimations
- Numerical aspect: Inverse and ill-posed problems
  - Applied mathematicians often are more interesting in existence, uniqueness, and construction, given an inifinite number of noise-free data, and stability given data contaminated by a deterministic disturbance.
  - Approach: *Tykhonov-Phillips* regularization and
  - numerical methods such as *L*-Curve (*Hansen* 1992) or the C_p-Plot (*Mallows* 1973).

2. The special linear *Gauss-Markov* model

Special Gauss Markov model

$$\mathbf{y} = \mathbf{A}\xi + \mathbf{e}$$

1st moments

$$\mathbf{A}\xi = E\{\mathbf{y}\},\ \mathbf{A} \in R^{n\times m},\ E\{\mathbf{y}\} \in \mathcal{R}(\mathbf{A}),\ \text{rk}\ \mathbf{A} = m \quad (1)$$

2nd moments

$$\boldsymbol{\Sigma}_{\mathbf{y}} = D\{\mathbf{y}\} \in R^{n\times n},\ \boldsymbol{\Sigma}_{\mathbf{y}} \text{ positive definite, rk } \boldsymbol{\Sigma}_{\mathbf{y}} = n \quad (2)$$

Theorem ($\hat{\xi}$ BLUUE of ξ) $\mathbf{y} = \mathbf{A}\xi + \mathbf{e}$

Let $\hat{\xi} = \mathbf{L}\mathbf{y}$ be $\boldsymbol{\Sigma}_{\mathbf{y}}$ - BLUUE $\hat{\xi}$ of ξ in the special linear Gauss-Markov model (1), (2) . Then

$$\mathrm{e}^T \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \mathrm{e} \rightarrow \min$$

$$\hat{\xi} = \mathbf{L}\mathbf{y} = (\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y}$$

$$\hat{\xi} = \boldsymbol{\Sigma}_{\hat{\xi}}\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y}$$

subject to the related dispersion matrix

$$D\{\hat{\xi}\} := \boldsymbol{\Sigma}_{\hat{\xi}} = (\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A}).^{-1}$$

$$MSE\{\hat{\xi}\} := E\{(\hat{\xi}-\xi)(\hat{\xi}-\xi)'\} = D\{\hat{\xi}\}$$

Best Linear Estimation (BLE) (R. Rao, 1972)

For the GGM model ($\mathbf{y}$, $\mathbf{A}\boldsymbol{\xi}$, $\Sigma_{\mathbf{y}} = \sigma^2 \mathbf{P}^{-1}$), and let $\mathbf{L}'\mathbf{y}$ be an estimator of $\boldsymbol{\xi}$; The *Mean Square Error* (MSE) of $\mathbf{L}'\mathbf{y}$ is :

$$E(\mathbf{L}'\mathbf{y}-\boldsymbol{\xi})^2 = \mathbf{L}'\Sigma_{\mathbf{y}}\mathbf{L} + (\mathbf{A}'\mathbf{L}-\mathbf{I}_m)'\boldsymbol{\xi}\boldsymbol{\xi}'(\mathbf{A}'\mathbf{L}-\mathbf{I}_m)$$

It is not a suitable criterion for minimizing, since it involves both the unknowns.

Three possibilities:

- Choose an a priori value of $\sigma^{-1}\boldsymbol{\xi}$, say $\boldsymbol{\gamma}$ and substitute $\sigma^2\mathbf{V} = \sigma^2 \boldsymbol{\gamma}\boldsymbol{\gamma}'$ for $\boldsymbol{\xi}\boldsymbol{\xi}'$;

$$F = \mathbf{L}'\mathbf{P}^{-1}\mathbf{L} + (\mathbf{A}'\mathbf{L}-\mathbf{I}_m)'\mathbf{V}(\mathbf{A}'\mathbf{L}-\mathbf{I}_m)$$

- If $\boldsymbol{\xi}$ is considered as a random variable with a priori mean dispersion $E(\boldsymbol{\xi}\boldsymbol{\xi}') = \sigma^2\mathbf{V}$;
- be consist of two parts, variance and bias. The choise of $\mathbf{V}$ in F represents the relative weight, which is n.n.d and of rank greater than one.
- yields the BLE of $\boldsymbol{\xi}$

$$\tilde{\boldsymbol{\xi}} = \mathbf{R}\mathbf{A}^T(\mathbf{P}^{-1} + \mathbf{A}\mathbf{V}\mathbf{A}^T)\mathbf{y} = (\mathbf{A}^T\mathbf{P}^{-1}\mathbf{A} + \mathbf{V}^{-1})\mathbf{A}^T\mathbf{P}^{-1}\mathbf{y}$$

■ α-weighted hybrid minimum variance- minimum bias estimation (hom α-BLE)

■ **The open problem to evaluate the regularization parameter**

➧ **Ever since *Tykhonov* (1963) and *Phillips* (1962) introduced the *hybrid minimum norm approximation solution* (HAPS) of a *linear improperly posed problem* there has been left the open problem to evaluate the regularization factor λ;**

➧ **In most applications of *Tykhonov-Phillips* type of regularization the weighting factor λ is determined by heurishical methods, such as by means of *L*-Curve (*Hansen* 1992) or the C_p-Plot (*Mallows* 1973). In literature also optimization techniques have been applied.**

α-weighted S-homBLE and A-optimal design of the regularization parameter λ

According to Grafarend and Schaffrin (1993), updated by Cai (2004), a homogeneously linear α-weighted hybrid minimum variance-minimum bias estimation (α, S-homBLE) is based upon the weighted sum of two norms of type:

$$\| MSE_{\alpha,\mathbf{S}}\{\hat{\boldsymbol{\xi}}\} \|^2 := \mathrm{tr}\,\mathbf{L}D\{\mathbf{y}\}\mathbf{L}' + \mathrm{tr}[\mathbf{I}_m - \mathbf{LA}]\frac{1}{\alpha}\mathbf{S}[\mathbf{I}_m - \mathbf{LA}]'$$

$$= \| \mathbf{L}' \|^2_{\boldsymbol{\Sigma}_\mathbf{y}} + \frac{1}{\alpha} \| (\mathbf{I}_m - \mathbf{LA})' \|^2_\mathbf{S}$$

namely

$$\textit{average variance} \quad \|\mathbf{L}'\|^2_{\boldsymbol{\Sigma}_\mathbf{y}} = \mathrm{tr}\,\mathbf{L}\boldsymbol{\Sigma}_\mathbf{y}\mathbf{L}'$$

$$\alpha, \mathbf{S} - \textit{weighted average bias} \quad \frac{1}{\alpha}\|(\mathbf{I}_m - \mathbf{LA})'\|^2_\mathbf{S} = \frac{1}{\alpha}\mathrm{tr}[\mathbf{I}_m - \mathbf{LA}]\mathbf{S}[\mathbf{I}_m - \mathbf{LA}]'.$$

- The hybrid norm $\| MSE_{\alpha,\mathbf{S}}\{\hat{\boldsymbol{\xi}}\} \|^2$ establishes the Lagrangean

$$\mathcal{L}(\mathbf{L}) := \| MSE_{\alpha,\mathbf{S}}\{\hat{\mathbf{L}}\} \|^2 = \mathrm{tr}\,\mathbf{L}\boldsymbol{\Sigma}_\mathbf{y}\mathbf{L}' + \frac{1}{\alpha}\mathrm{tr}\,(\mathbf{I}_m - \mathbf{LA})\mathbf{S}(\mathbf{I}_m - \mathbf{LA})' = \min_\mathbf{L}$$

for $\hat{\boldsymbol{\xi}}$ as α,S-homBLE of $\boldsymbol{\xi}$ (*Theorem* 1).

Theorem 1 $\hat{\xi}$ α,S-homBLE, also called: *ridge estimator*

***Linear Gauss-Markov model*:**

$$\mathbf{A}\xi = E\{\mathbf{y}\},\ \mathbf{A} \in R^{n\times m},\ E\{\mathbf{y}\} \in \mathcal{R}(\mathbf{A}),\ \text{rk}\ \mathbf{A} = m$$
$$\boldsymbol{\Sigma}_\mathbf{y} = D\{\mathbf{y}\} \in R^{n\times n},\ \boldsymbol{\Sigma}_\mathbf{y}\ \text{positive definite, rk}\ \boldsymbol{\Sigma}_\mathbf{y} = n$$

***α,S-homBLE*:**

$$\hat{\xi} = (\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{y}\quad (\text{if}\ \ \mathbf{S}^{-1}\ \text{exists})$$

***dispersion matrix* :**

$$D\{\hat{\xi}\} = (\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}$$

***bias vector*:**

$$\boldsymbol{\beta} := E\{\hat{\xi}\} - \xi = -[\mathbf{I}_m - (\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A}]\xi =$$
$$= -\alpha(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{S}^{-1}\xi$$

***Mean Square Error matrix*:**

$$MSE\{\hat{\xi}\} := E\{(\hat{\xi} - \xi)(\hat{\xi} - \xi)'\} = D\{\hat{\xi}\} + \boldsymbol{\beta}\boldsymbol{\beta}' =$$
$$= (\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1} +$$
$$+[(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}\alpha\mathbf{S}^{-1}]\xi\xi'[\alpha\mathbf{S}^{-1}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}] =$$
$$= (\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}[\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + (\alpha\mathbf{S}^{-1})\xi\xi'(\alpha\mathbf{S}^{-1})](\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}.$$

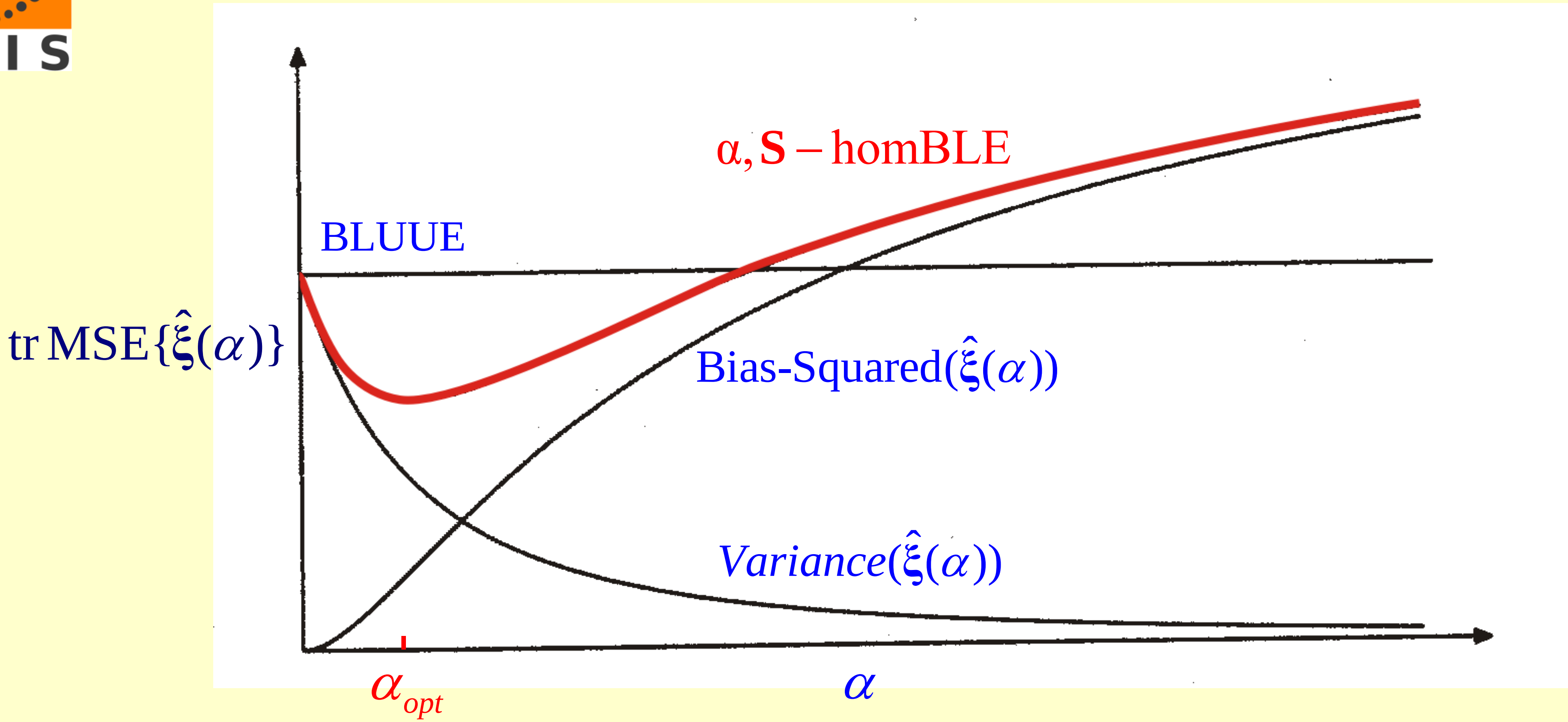

Figure 1. The relationship between the variance, the squared bias and the weighting factor α. The variacne term decrease as α increases, while the squared bias increase with α.

The geodetic inverse Problem:

- **Exact or strict Multicollinearity means** $|\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}\mathbf{A}| = 0$
- **weak Multicollinearity means** $|\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}\mathbf{A}| \approx 0$
- **Use the condition Number for diagnostics:** $k = \left(\frac{\lambda_{\max}}{\lambda_{\min}}\right)^{\frac{1}{2}}$

The weight factor α can be alternatively determined by the *A-optimal design* of type

$$(1) \quad \operatorname{tr} D\{\hat{\boldsymbol{\xi}}\} = \min_{\alpha}, \text{ or}$$

$$(2) \quad \boldsymbol{\beta}\boldsymbol{\beta}' = \min_{\alpha}, \text{ or}$$

$$(3) \quad \operatorname{tr} \mathrm{MSE}\{\hat{\boldsymbol{\xi}}\} = \min_{\alpha}$$

Here we focus on the *third case* – the most meaningful one –

"minimize the trace of the *Mean Square Error matrix* $\operatorname{tr}\mathrm{MSE}\{\hat{\boldsymbol{\xi}}\}$ **of** $\hat{\boldsymbol{\xi}}$**-weighted S-homBLE** (α) **to find**

$$\hat{\alpha} = \arg\{\operatorname{tr}\mathrm{MSE}\{\hat{\boldsymbol{\xi}}\} = \min\}"$$

Theorem 2. A-optimal design of α

Let the average hybrid α-weighted variance-bias norm $\mathrm{MSE}\{\hat{\boldsymbol{\xi}}\}$ **of** $\hat{\boldsymbol{\xi}}$ **(α, S-homBLE) with respect to the linear *Gauss-Markov* model be given by**

$$\operatorname{tr}\mathrm{MSE}\{\hat{\boldsymbol{\xi}}\} =$$

$$= \operatorname{tr}(\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1} +$$

$$+ \operatorname{tr}[(\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}\alpha\mathbf{S}^{-1}\boldsymbol{\xi}\boldsymbol{\xi}'[\alpha\mathbf{S}^{-1}(\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}],$$

then follows by A-optimal design in the sense of

$$\operatorname{tr}\mathrm{MSE}\{\hat{\boldsymbol{\xi}}\} = \min$$

if and only if

$$\hat{\alpha} = \frac{\operatorname{tr}\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A} + \hat{\alpha}\mathbf{S}^{-1})^{-2}\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}}{\boldsymbol{\xi}'\mathbf{S}^{-1}(\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A} + \hat{\alpha}\mathbf{S}^{-1})^{-2}\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{S}^{-1}\boldsymbol{\xi}}$$

Tykhonov-Phillips regularization:

Tykhonov-Phillips regularization is defined as the solution to the problem

$$\min_{\xi}\{\|\mathbf{A}\xi-\mathbf{y}\|_{\mathbf{W}}^{2}+\alpha\|\xi\|^{2}\}.$$

yields the normal equations system

$$\tilde{\xi}=(\mathbf{A}^{T}\mathbf{W}\mathbf{A}+\alpha\,\mathbf{I})^{-1}\mathbf{A}^{T}\mathbf{W}\mathbf{y}.$$

Ridge regression (Hoerl and Kennard, 1970a, b).

Ridge regression is defined as biased estimation for nonorthogonal problems with a *Lagrangian* function:

$$\text{minimize}\;\; F=\xi'\xi+(1/k)[(\xi-\hat{\xi})^{T}\mathbf{A}^{T}\mathbf{A}(\xi-\hat{\xi})-\phi_{0}].$$

or a equivalent statement: $F_{1}=(\mathbf{y}-\mathbf{A}\xi)^{T}(\mathbf{y}-\mathbf{A}\xi)+k(\xi'\xi-R^{2}).$

yields the normal equations system

$$\tilde{\xi}=(\mathbf{A}^{T}\mathbf{A}+k\,\mathbf{I})^{-1}\mathbf{A}^{T}\mathbf{y}=[\mathbf{I}_{p}+k(\mathbf{A}^{T}\mathbf{A})^{-1}]^{-1}\hat{\xi}.$$

Generalized Tykhonov-Phillips regularization:

Tykhonov-Phillips regularization is defined as the solution to the problem

$$\min_{\xi}\{\|\mathbf{A}\xi-\mathbf{y}\|_{\mathbf{W}}^2+\|\xi-\xi_0\|_{\mathbf{R}}^2\}.$$

What is the solution of generalized regularization ?

$$\tilde{\xi}=(\mathbf{A}^T\mathbf{W}\mathbf{A}+\mathbf{R})^{-1}\mathbf{A}^T\mathbf{W}\mathbf{y}.$$

Rewrite the objective function

$$\min_{\xi}\{\|\mathbf{A}(\xi-\xi_0)-(\mathbf{y}-\mathbf{A}\xi_0)\|_{\mathbf{W}}^2+\|\xi-\xi_0\|_{\mathbf{R}}^2\}.$$

yields the right solution

$$\begin{aligned}
\tilde{\xi} &= \xi_0+(\mathbf{A}^T\mathbf{W}\mathbf{A}+\mathbf{R})^{-1}\mathbf{A}^T\mathbf{W}(\mathbf{y}-\mathbf{A}\xi_0)= \\
&= (\mathbf{A}^T\mathbf{W}\mathbf{A}+\mathbf{R})^{-1}\mathbf{A}^T\mathbf{W}\mathbf{y}-(\mathbf{I}-(\mathbf{A}^T\mathbf{W}\mathbf{A}+\mathbf{R})^{-1}\mathbf{A}^T\mathbf{W}\mathbf{A})\xi_0 \\
&= (\mathbf{A}^T\mathbf{W}\mathbf{A}+\mathbf{R})^{-1}\mathbf{A}^T\mathbf{W}\mathbf{y}+(\mathbf{A}^T\mathbf{W}\mathbf{A}+\mathbf{R})^{-1}\mathbf{R}\xi_0 \\
&= (\mathbf{A}^T\mathbf{W}\mathbf{A}+\mathbf{R})^{-1}\mathbf{A}^T\mathbf{W}\mathbf{y}+(\mathbf{A}^T\mathbf{W}\mathbf{A}+\mathbf{R})^{-1}\mathbf{R}\xi_0 \\
&= (\mathbf{A}^T\mathbf{W}\mathbf{A}+\mathbf{R})^{-1}\mathbf{A}^T\mathbf{W}\mathbf{y}+(\mathbf{A}^T\mathbf{W}\mathbf{A}+\mathbf{R})^{-1}\mathbf{R}\xi_0 \\
&= (\mathbf{A}^T\mathbf{W}\mathbf{A}+\mathbf{R})^{-1}(\mathbf{A}^T\mathbf{W}\mathbf{y}+\mathbf{R}\xi_0)
\end{aligned}$$

■ Comparison of the determination of the regularization factor λ by A-optimal design and the ridge parameter *k* in ridge regression

- ***Hoerl, Kennard and Baldwin*** **(1975) have suggested that if** $\mathbf{A'A=I_m}$**, then a minimum meansquare error (MSE) is obtained if *ridge parameter*** $k = m\sigma^2/\xi'\xi$ **for multiple linear regression model:**

$$\mathbf{y}=\mathbf{A}\boldsymbol{\xi}+\mathbf{e},\ \mathbf{A}\in\mathbb{R}^{n\times m},\ \mathrm{rk}\,\mathbf{A}=m,\ E\{\mathbf{e}\}=\mathbf{0},\ D\{\mathbf{y}\}=\mathbf{I}_n\sigma^2,\ \sigma^2\ \text{unknown}$$

- **This is just the special case of our general solution by A-optimal design of Corollary 3 under unit weight P and A’A=I$_{\mathbf{m}}$, yielding**

$$\hat{\lambda}=\frac{\mathrm{tr}\,\mathbf{A'A}(\mathbf{A'A}+\hat{\lambda}\mathbf{I}_m)^{-3}\sigma^2}{\boldsymbol{\xi}'(\mathbf{A'A}+\hat{\lambda}\mathbf{I}_m)^{-2}\mathbf{A'A}(\mathbf{A'A}+\hat{\lambda}\mathbf{I}_m)^{-1}\boldsymbol{\xi}}=\frac{\mathrm{tr}\,\mathbf{I}_m(\mathbf{I}_m+\lambda\mathbf{I}_m)^{-3}\sigma^2}{\boldsymbol{\xi}'(\mathbf{I}_m+\lambda\mathbf{I}_m)^{-2}\mathbf{I}_m(\mathbf{I}_m+\lambda\mathbf{I}_m)^{-1}\boldsymbol{\xi}}=$$

$$=\frac{m(1+\hat{\lambda})^{-3}\sigma^2}{\boldsymbol{\xi}'\boldsymbol{\xi}(1+\hat{\lambda})^{-3}}=\frac{m\sigma^2}{\boldsymbol{\xi}'\boldsymbol{\xi}}.$$

3. Mixed model (Combination method)

Theil & Goldberger (1961) and *Theil* (1963), *Toutenburg* (1982) and *Rao & Toutenburg* (1999) *mixed estimator* with additional information as stochastic linear restrictions:

***Linear Gauss-Markov model*:**

$$\mathbf{y} = \mathbf{A}\xi + \mathbf{e},\quad E\{\mathbf{e}\} = \mathbf{0},\quad D\{\mathbf{e}\} = \boldsymbol{\Sigma} \in \mathbb{R}^{n\times n},$$
$$\mathbf{A} \in \mathbb{R}^{n\times m},\ \text{rk}\,\mathbf{A} = m,\ \text{rk}\,\boldsymbol{\Sigma} = n.$$

The additional information:

$$\xi_p,\ E\{\xi_p\} = E\{\hat{\xi}\} = E\{\xi\}$$

Stochastic linear restriction:

$$\xi_P = \mathbf{I}\xi + \mathbf{e}_P,\quad E\{\mathbf{e}_P\} = \mathbf{0},\ D\{\mathbf{e}_P\} = \boldsymbol{\Sigma}_P.$$

***Mixed model*:**

$$\begin{bmatrix}\mathbf{y}\\ \xi_P\end{bmatrix} = \begin{bmatrix}\mathbf{A}\\ \mathbf{I}\end{bmatrix}\xi + \begin{bmatrix}\mathbf{e}\\ \mathbf{e}_P\end{bmatrix},\quad E\left\{\begin{bmatrix}\mathbf{e}\\ \mathbf{e}_P\end{bmatrix}\right\} = \mathbf{0},\ D\left\{\begin{bmatrix}\mathbf{e}\\ \mathbf{e}_P\end{bmatrix}\right\} = \begin{bmatrix}\boldsymbol{\Sigma} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma}_P\end{bmatrix}.$$

The BLUUE estimator of the *original G-M model*:

$$\hat{\xi}_0 = (\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{y}.$$

The BLUUE estimator of the *mixed model*:

$$\hat{\xi}(P) = (\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A} + \boldsymbol{\Sigma}_P^{-1})^{-1}(\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{y} + \boldsymbol{\Sigma}_P^{-1}\xi_p)$$
$$= \hat{\xi}_0 + (\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}[\boldsymbol{\Sigma}_P + (\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}]^{-1}(\xi_P - \hat{\xi}_0).$$

The dispersion matrix:

$$\boldsymbol{\Sigma}_{\hat{\xi}(P)} = (\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A} + \boldsymbol{\Sigma}_P^{-1})^{-1}$$

$\hat{\xi}(P)$ – **unbiased with smaller dispersion:**

$$\boldsymbol{\Sigma}_{\hat{\xi}_0} - \boldsymbol{\Sigma}_{\hat{\xi}(P)} = (\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1} - (\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A} + \boldsymbol{\Sigma}_P^{-1})^{-1} =$$
$$= (\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}[\boldsymbol{\Sigma}_P + (\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}]^{-1}(\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1} \geq 0$$

➡ **The use of stochastic restrictions leads to a gain in efficiency.**

The light constraint solutions (*Reigber*, 1989)

Light constraint with a priori information $\xi - \xi_p = 0$

$$0 = -\mathbf{I}\xi + \xi_p,\quad \xi_p \sim (\mathbf{0}, \Sigma_P) \quad i.e.\ E\{\xi_p\} = \mathbf{0},\ D\{\xi_p\} = \Sigma_P .$$

Light constraint model :

$$\begin{bmatrix}\mathbf{y}\\ \mathbf{0}\end{bmatrix} = \begin{bmatrix}\mathbf{A}\\ \mathbf{I}\end{bmatrix}\xi + \begin{bmatrix}\mathbf{e}\\ \xi_p\end{bmatrix},\ E\left\{\begin{bmatrix}\mathbf{e}\\ \xi_p\end{bmatrix}\right\} = \mathbf{0},\ D\left\{\begin{bmatrix}\mathbf{e}\\ \xi_p\end{bmatrix}\right\} = \begin{bmatrix}\Sigma & \mathbf{0}\\ \mathbf{0} & \Sigma_P\end{bmatrix}.$$

BLUUE estimator of ***light constraint model :***

$$\begin{aligned}\hat{\xi}(L) &= (\mathbf{A}'\Sigma^{-1}\mathbf{A} + \Sigma_P^{-1})^{-1}(\mathbf{A}'\Sigma^{-1}\mathbf{y} + \Sigma_P^{-1}\mathbf{0})\\ &= (\mathbf{A}'\Sigma^{-1}\mathbf{A} + \Sigma_P^{-1})^{-1}\mathbf{A}'\Sigma^{-1}\mathbf{y}\\ &= \hat{\xi}_0 - (\mathbf{A}'\Sigma^{-1}\mathbf{A})^{-1}[\Sigma_P + (\mathbf{A}'\Sigma^{-1}\mathbf{A})^{-1}]^{-1}\hat{\xi}_0 .\end{aligned}$$

With the dispersion matrix: $\Sigma_{\hat{\xi}(L)} = (\mathbf{A}'\Sigma^{-1}\mathbf{A} + \Sigma_P^{-1})^{-1}$.

The objective function of light constraint solution:

$$\begin{bmatrix}\mathbf{e}' & \xi_p'\end{bmatrix}\begin{bmatrix}\Sigma & 0\\ 0 & \Sigma_P\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{e}\\ \xi_p\end{bmatrix} = \mathbf{e}'\Sigma^{-1}\mathbf{e} + \xi_p'\Sigma_P^{-1}\xi_p ,$$

- **The same as the objective function of the estimate with weighting parameters !**

Difference between estimators of the *mixed models* and *light constraint models*:

$$\begin{aligned}\Delta\hat{\xi} = \hat{\xi}(P) - \hat{\xi}(L) &= (\mathbf{A}'\Sigma^{-1}\mathbf{A} + \Sigma_P^{-1})^{-1}(\mathbf{A}'\Sigma^{-1}\mathbf{y} + \Sigma_P^{-1}\xi_p) - (\mathbf{A}'\Sigma^{-1}\mathbf{A} + \Sigma_P^{-1})^{-1}\mathbf{A}'\Sigma^{-1}\mathbf{y}\\ &= (\mathbf{A}'\Sigma^{-1}\mathbf{A} + \Sigma_P^{-1})^{-1}\Sigma_P^{-1}\xi_p .\end{aligned}$$

Combination and Regularization methods

In order to solve the PGP by spectral domain stabilization, we use a priori information in terms of spherical harmonic coefficients. Augmenting the minimization of squared residuals $\mathbf{r} = \mathbf{A}\boldsymbol{\xi} - \mathbf{y}$ by a parameter component $\boldsymbol{\xi} - \boldsymbol{\xi}_0$

$$\min_{\xi}\{\| \mathbf{A}\boldsymbol{\xi} - \mathbf{y} \|^2_{\boldsymbol{\Sigma}_\mathbf{y}^{-1}} + \alpha \| \boldsymbol{\xi} - \boldsymbol{\xi}_0 \|^2_{\mathbf{R}} \}.$$

yields the normal equations system

$$\hat{\boldsymbol{\xi}} = (\mathbf{A}^T \boldsymbol{\Sigma}_\mathbf{y}^{-1} \mathbf{A} + \alpha \mathbf{R})^{-1} (\mathbf{A}^T \boldsymbol{\Sigma}_\mathbf{y}^{-1} \mathbf{y} + \alpha \mathbf{R} \boldsymbol{\xi}_0) .$$

- The parameter α denotes the regularization parameter. It balances the residual norm $||\mathbf{A}\boldsymbol{\xi} - \mathbf{y}||$ against the (reduced) parameter norm $||\boldsymbol{\xi} - \boldsymbol{\xi}_0||$.
- Accounts for both regularization and combination, which is just the so-called **generalized Tykhonov regularization in the case of α = 1!**

• When $\boldsymbol{\xi}_0 = 0$, i.e. the a priori information consisting of null pseudo-observables, This is the case of regularization.

•Data combination in the spectral domain is achieved by incorporating non-trivial a priori information $\boldsymbol{\xi}_0 \neq 0$, yielding the mixed estimator with additional information as stochastic linear restrictions.

Biased and unbiased estimations in different aspects

Numerical Analysis | **Statistical aspects**

Standard Tykhonov-Phillips regularization

$$\min_{\xi}\{\|\mathbf{A}\xi-\mathbf{y}\|^2_{\mathbf{W}}+\|\xi\|^2_{\mathbf{R}}\}.$$

$$\hat{\xi}=(\mathbf{A}^T\mathbf{W}\mathbf{A}+\mathbf{R})^{-1}\mathbf{A}^T\mathbf{W}\mathbf{y}.$$

Ridge regression, BLE and Light constraint solution

$$\|MSE\{\hat{\xi}\}\|^2:=\|E\{(\hat{\xi}-\xi)(\hat{\xi}-\xi)'\}\|^2=\min$$

$$\hat{\xi}=(\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A}+\mathbf{V}^{-1})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y}.$$

Biased

Gen.Tykhonov-Phillips regularization

$$\min_{\xi}\{\|\mathbf{A}\xi-\mathbf{y}\|^2_{\mathbf{W}}+\|\xi-\xi_0\|^2_{\mathbf{R}}\}.$$

$$\tilde{\xi}=(\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}+\mathbf{R})^{-1}(\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{y}+\mathbf{R}\xi_0).$$

BLUUE estimator of Mixed model

$$\mathbf{e}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{e}+\xi_p'\boldsymbol{\Sigma}_P^{-1}\xi_p\rightarrow\min$$

$$\tilde{\xi}=(\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A}+\boldsymbol{\Sigma}_P^{-1})^{-1}(\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y}+\boldsymbol{\Sigma}_P^{-1}\xi_p)$$

Unbiased

- **This answer the relationship of these biased and unbiased solutions and estimators in numerical and statistical aspects.**

4. Conclusions and further studies

- **Development of a rigorous approach to minimum MSE adjustment in a Gauss-Markov Model, i.e. α -weighted S-homBLE;**
- **Derivation of a new method of determining the optimal regularization parameter α in uniform Tykhonov-Phillips regularization (α-weighted S-homBLE) by A-optimal design in the general case;**
- **It was, therefore, possible to translate the previous results for the α -weighted S-homBLE to the case of Tykhonov-Phillips regularization with remarkable success;**
- **The optimal ridge parameter *k* in *ridge regression* as developed by *Hoerl and Kennard* in 1970s is just the special case of our general solution by A-optimal design.**
- **Accounts for both regularization and combination, which is just the so-called generalized Tykhonov regularization!**

In order to develop and promote the generality of inversion methods, it is necessary to study this kind problem from the following aspects:

1) Statistical or deterministic regularization;

2) Ridge estimation;

3) Best linear estimation;

4) Mixed model;

5) Biased or unbiased estimations;

6) The criterion in derivation of the inversion solution:

Mean square error of the estimtes MSE - Gauss´s second approach

$$MSE\{\hat{\xi}\} := E\{(\hat{\xi}-\xi)(\hat{\xi}-\xi)'\} = D\{\hat{\xi}\} + \boldsymbol{\beta}\boldsymbol{\beta}'$$

instead of Gauss´s first approach $\mathrm{e}^T \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \mathrm{e} \rightarrow \min$

7) Optimal solution.

Historical remark:

Laplace (1810) distinguishes between errors of observaions and errors of estimates, and points out that a theory of estimation should be based on a measure of derviation between the estimate and the true value.

Gauss(1823) finally accepted Laplace's criticism and indicates that if he were to rewrite the 1809 proof (LS), he would use the expected mean square error as optimality criterion.

This means that estimation theory should be based on minimization of the error of estimation!

Journal of the American Statistical Association, Vol. 63, No. 321 (Mar., 1968)

SOME SHRINKAGE TECHNIQUES FOR ESTIMATING THE MEAN

JAMES R. THOMPSON
Vanderbilt University

1. INTRODUCTION

IN A paper presented to the Royal Society of Göttingen in 1809, Karl Friedrich Gauss [3] proposed that the smallness of the mean square error of an estimator about its estimand be employed as a measure of its excellence:

> "From the value of the integral $\int_{-\infty}^{\infty} x\phi(x)dx$, i.e. the average value of x, we learn the existence or nonexistence of a constant error as well as the value of this error; similarly, the integral $\int_{-\infty}^{\infty} x^2\phi(x)dx$, i.e. the average value of x^2, seems very suitable for defining and measuring, in a general way, the uncertainty of a system of observations. . . . If one objects that this convention is arbitrary and does not appear necessary, we readily agree. The question which concerns us here has something vague about it from its very nature, and cannot be made really precise except by some principle which is arbitrary to a certain degree. . . . It is clear to begin with that the loss should not be proportional to the error committed, for under this hypothesis, since a positive error represents a loss, a negative error would be considered as a gain; the magnitude of the loss ought, on the contrary, to be evaluated by a function of the error whose value is always positive. Among the infinite number of functions satisfying this condition, it seems natural to choose the simplest, which is, without doubt, the square of the error, and in this way we are led to the principle proposed above."

In this and subsequent works, he made a thorough study of least-squares estimation, which was motivated by the MSE criterion. Nearly a century later, Markov [7] proved that the minimum variance unbiased linear estimator of the means of a set of independent random variables is the least-squares estimator.

[3] Gauss, Karl F., "Theoria Combinationis Observationum Erroribus Minimis Obnoxiae," *Werke*, Göttingen, 1821, Vol. 4, pp. 6–7 (*Gauss's Work on the Theory of Least Squares*, trans. by Hale F. Trotter, STRG report No. 5, Princeton University, 1957).

THEORIA

COMBINATIONIS OBSERVATIONUM

ERRORIBUS MINIMIS OBNOXIAE

PARS PRIOR

AUCTORE

CAROLO FRIDERICO GAUSS

SOCIETATI REGIAE SCIENTIARUM EXHIBITA 1821. FEBR. 15.

Commentationes societatis regiae scientiarum Gottingensis recentiores. Vol. v.
Gottingae MDCCCXXIII.

CARL FRIEDRICH GAUSS

WERKE

VIERTER BAND.

HERAUSGEGEBEN

VON DER

KÖNIGLICHEN GESELLSCHAFT DER WISSENSCHAFTEN

ZU

GÖTTINGEN

1873.

1821, 1823 und 1826: *Theoria combinationis observationum erroribus minimis obnoxiae*. Drei Abhandlungen betreffend die Wahrscheinlichkeitsrechnung als Grundlage des Gauß'schen Fehlerfortpflanzungsgesetzes. English translation by G. W. Stewart, 1987, Society for Industrial Mathematics.

6.

Perinde ut integrale $\int x\varphi x . dx$, seu valor medius ipsius x, erroris constantis vel absentiam vel praesentiam et magnitudinem docet, integrale

$$\int xx\varphi x . dx$$

ab $x = -\infty$ usque ad $x = +\infty$ extensum (seu valor medius quadrati xx) aptissimum videtur ad incertitudinem observationum in genere definiendam et dimetiendam, ita ut e duobus observationum systematibus, quae quoad errorum facilitatem inter se differunt, eae praecisione praestare censeantur, in quibus integrale $\int xx\varphi x . dx$ valorem minorem obtinet. Quodsi quis hanc rationem pro arbitrio, nulla cogente necessitate, electam esse obiiciat, lubenter assentiemur.

Theoria Combinationis Observationum Erroribus Minimis Obnoxiae

Pars Prior ◆ Pars Posterior ◆ Supplementum

By Carl Friedrich Gauss

Theory of the Combination of Observations Least Subject to Errors

Part One ◆ Part Two ◆ Supplement

Translated by G. W. Stewart
University of Maryland

siam
Society for Industrial and Applied Mathematics
Philadelphia 1995

- ***Thank you !***