Weierstraß-Institut für Angewandte Analysis und Stochastik

Vladimir Spokoiny

# Saddle Point Model Selection in Inverse Problems

## Outline

## Outline

## Model

- Observed a vector $\boldsymbol{Y} = (Y_1, \ldots, Y_n) \in \mathbb{R}^n$ .
- Model

$$Y_i = s_i + \sigma_i \varepsilon_i$$

where $\varepsilon_i$ are iid standard normal and $\sigma_i$ are given numbers.

- Direct problem: $\sigma_i \equiv \sigma$ .
- Inverse problem: $\sigma_i \nearrow \infty$ .
- Target: $\boldsymbol{s} = (s_i)$ .

## Linear estimates

Linear estimates $\widetilde{\boldsymbol{s}}$ of $\boldsymbol{s}$ defined by a vector $\varkappa$

$$\widetilde{\boldsymbol{s}} = \widetilde{\boldsymbol{s}}(\varkappa).$$

The estimation loss and risk are given by

$$\rho(\varkappa) = \sum_i \bigl[\widetilde{s}_i(\varkappa) - s_i\bigr]^2,$$

$$\mathcal{R}(\varkappa) = I\!\!E \sum_i \bigl[\widetilde{s}_i(\varkappa) - s_i\bigr]^2.$$

Oracle choice of an estimate:

$$\varkappa^* \stackrel{\text{def}}{=} \operatorname*{argmin}_{\varkappa} \mathcal{R}(\varkappa).$$

## Some previous works

- Unbiased Risk Estimation (SURE): $C_p$-Mallows (73), Akaike (74), Donoho and Johnstone (1995).
- Blockwise SURE: Pinsker (1984), Tsybakov (2001),
- Adaptive estimation for inverse problems: Cavalier and Tsybakov (2002), Goldenshluger (1999), Cavalier, Golubev, Lepski, Tsybakov (2003).
- Penalized Gaussian Model Selection (PMS): Birgé and Massart (2001, 2007), Baraud, Huet and Laurent (2003).
- Pairwise comparison, (Test) T-estimation: Lepski (1990), Goldenshluger and Lepski (2009), Birgé (2006), Baraud et al. (2009).
- Sparse Methods: thresholding, LASSO, regularized LASSO etc.: Donoho and Johnstone (1995), Tibshirani (1995), Candes Tao (2007), van de Geer (2007), Bickel, Ritov, Tsybakov (2009) . . .

## Unbiased risk estimate

The unbiased risk estimate is given by

$$\ell(\varkappa) = \sum_i \Bigl\{ \bigl[\widetilde{s}_i(\varkappa) - Y_i\bigr]^2 - \mathbb{E}\bigl[\widetilde{s}_i(\varkappa) - Y_i\bigr]^2 + \mathbb{E}\bigl[\widetilde{s}_i(\varkappa) - s_i\bigr]^2 \Bigr\},$$

where

$$\mathbb{E}\bigl[\widetilde{s}_i(\varkappa) - Y_i\bigr]^2 - \mathbb{E}\bigl[\widetilde{s}_i(\varkappa) - s_i\bigr]^2$$

does not depend on $\boldsymbol{s}$ because $\widetilde{\boldsymbol{s}}(\varkappa)$ is linear. Obviously,

$$\mathbb{E}\ell(\varkappa) = \mathbb{E}\sum_i \bigl[\widetilde{s}_i(\varkappa) - s_i\bigr]^2 = \mathcal{R}(\varkappa).$$

## SURE/AIC method

The minimization of SURE: define $\widetilde{\varkappa}$ by minimizing the empirical counterpart $\ell(\varkappa)$ of $\mathcal{R}(\varkappa)$ :

$$\widetilde{\varkappa} = \operatorname*{argmin}_{\varkappa} \ell(\varkappa).$$

Aim: to mimic the oracle choice:

$$\varkappa^* \stackrel{\text{def}}{=} \operatorname*{argmin}_{\varkappa} \mathcal{R}(\varkappa).$$

Problem: poor performance for inverse problems, especially for severely ill-posed.

## PMS

Ultimate Goal: establish a concentration property (oracle bound) regardless particular features of the risk function and degree of ill-posedness.

Penalized model selection (PMS) approach:

$$\widehat{\varkappa} = \operatorname*{argmin}_{\varkappa} \bigl\{ \ell(\varkappa) + 2\,\mathrm{pen}(\varkappa) \bigr\},$$

where $z(\varkappa)$ mimics the complexity related to the choice of the estimate $\widetilde{\boldsymbol{s}}(\varkappa)$. Can be rewritten in the form

$$\ell(\widehat{\varkappa}) + 2\,\mathrm{pen}(\widehat{\varkappa}) \le \ell(\varkappa) + 2\,\mathrm{pen}(\varkappa)$$

for all $\varkappa$, or equivalently,

$$\ell(\widehat{\varkappa}, \varkappa) \le 2\,\mathrm{pen}(\widehat{\varkappa}) - 2\,\mathrm{pen}(\varkappa).$$

## Limitations of PMS. Alternative approaches

Major problem: $2\bigl[\mathrm{pen}(\varkappa') - \mathrm{pen}(\varkappa)\bigr]$ is a poor critical value for $\ell(\varkappa', \varkappa)$.

Alternative idea (to PMS) is to consider a bivariate function
$\ell(\varkappa', \varkappa) = \ell(\varkappa') - \ell(\varkappa)$.

Related proposals:

- Lepski (1990): ordered model selection via pairwise comparison of different estimates. Extensions to inverse problems due to Mathe and Pereversev among others.
- Goldenshluger and Lepski (2009): extension to unordered case.
- Birgé (2006): T-estimation (ideas already in Le Cam 1973, Birgé 1983).
- Dümbgen and Rohde (2009): for model selection via multiple testing; worst case calibration.

## Risk minimization vs saddle point

Consider the risk difference as functions of two arguments $\varkappa, \varkappa'$ :

$$\mathcal{R}(\varkappa, \varkappa') \stackrel{\text{def}}{=} \mathcal{R}(\varkappa) - \mathcal{R}(\varkappa').$$

The definition of $\varkappa^*$ as the minimizer of the risk function $\mathcal{R}(\varkappa)$ can be rewritten in the form:

$$\varkappa^* = \operatorname*{argmin}_{\varkappa} \mathcal{R}(\varkappa, \varkappa^*) = \operatorname*{argmax}_{\varkappa} \mathcal{R}(\varkappa^*, \varkappa).$$

In other words, $(\varkappa^*, \varkappa^*)$ is a saddle point of the function $\mathcal{R}(\varkappa, \varkappa')$ :

$$\varkappa^* = \operatorname*{argmin}_{\varkappa} \max_{\varkappa'} \mathcal{R}(\varkappa, \varkappa') = \operatorname*{argmax}_{\varkappa} \min_{\varkappa'} \mathcal{R}(\varkappa, \varkappa')$$

## Model selection as a SP-problem

Basic observation: the minimizer of SURE $\ell(\varkappa)$ is a saddle point of $\ell(\varkappa, \varkappa') = \ell(\varkappa) - \ell(\varkappa')$.

SP-Approach: consider a penalized bivariate function

$$\ell(\varkappa, \varkappa') + 2\,\mathrm{PEN}(\varkappa, \varkappa') = \ell(\varkappa) - \ell(\varkappa') + 2\,\mathrm{PEN}(\varkappa, \varkappa')$$

and define the penalized choice of $\widehat{\varkappa}$ via the smallest saddle point:

$$\widehat{\varkappa} \stackrel{\mathrm{def}}{=} \operatorname*{argmin}_{\varkappa} \max_{\varkappa'} \{\ell(\varkappa, \varkappa') - 2\,\mathrm{PEN}(\varkappa, \varkappa')\}.$$

In other words, the value $\widehat{\varkappa}$ is a SP if it holds for any $\varkappa^{\circ}$:

$$\max_{\varkappa} \{\ell(\widehat{\varkappa}, \varkappa) - 2\,\mathrm{PEN}(\widehat{\varkappa}, \varkappa)\} \le \max_{\varkappa} \{\ell(\varkappa^{\circ}, \varkappa) - 2\,\mathrm{PEN}(\varkappa^{\circ}, \varkappa)\}.$$

## Implication of the SP choice

The value $\hat{\varkappa}$ is a SP if it holds for any $\varkappa^{\circ}$:

$$\max_{\varkappa}\{\ell(\hat{\varkappa},\varkappa) - 2\,\mathrm{PEN}(\hat{\varkappa},\varkappa)\} \le \max_{\varkappa}\{\ell(\varkappa^{\circ},\varkappa) - 2\,\mathrm{PEN}(\varkappa^{\circ},\varkappa)\}.$$

In particular, this inequality can be applied with $\varkappa = \varkappa^{*}$ yielding

$$\ell(\hat{\varkappa},\varkappa^{*}) - 2\,\mathrm{PEN}(\hat{\varkappa},\varkappa^{*}) \le \max_{\varkappa}\{\ell(\varkappa^{*},\varkappa) - 2\,\mathrm{PEN}(\varkappa^{*},\varkappa)\}. \qquad (1)$$

If the penalty $\mathrm{PEN}(\varkappa,\varkappa')$ ensures (with high proba) that

$$\max_{\varkappa}\{\ell(\varkappa^{*},\varkappa) - 2\,\mathrm{PEN}(\varkappa^{*},\varkappa)\} \le \xi^{*}$$

Then (1) implies

$$\boxed{\ell(\hat{\varkappa},\varkappa^{*}) - 2\,\mathrm{PEN}(\hat{\varkappa},\varkappa^{*}) \le \xi^{*}.}$$

## Relation to multiple testing idea

"$\widehat{\varkappa}$ is a SP" can be rewritten as

$$\ell(\widehat{\varkappa}) - \ell(\varkappa) \le t(\widehat{\varkappa}, \varkappa)$$

for all $\varkappa > \widehat{\varkappa}$, where

$$t(\widehat{\varkappa}, \varkappa) = 2\,\mathrm{PEN}(\varkappa, \widehat{\varkappa}).$$

Interpretation: $T(\varkappa, \varkappa') = \ell(\varkappa) - \ell(\varkappa')$, a test statistic for testing $\varkappa$ against $\varkappa'$ and $t(\varkappa, \varkappa')$ is the critical value of the test.

$\widehat{\varkappa}$ accepted if $\widehat{\varkappa}$ is not rejected against any $\varkappa > \widehat{\varkappa}$.

## Key words for the next:

- ▷ Cross-concentration bounds for excess loss
- ▷ One-sided concentration
- ▷ Oracle bound for saddle point model selection

## Outline

## SURE: Problem formulation

“Oracle”:

$$\varkappa^* \stackrel{\text{def}}{=} \operatorname*{argmin}_{\varkappa} \mathcal{R}(\varkappa)$$

Define

$$L_1(\varkappa, \varkappa^*) \stackrel{\text{def}}{=} \frac{1}{2}\bigl\{\ell(\varkappa^*) - \ell(\varkappa)\bigr\}.$$

Then

$$\widetilde{\varkappa} = \operatorname*{argmin}_{\varkappa} \ell(\varkappa) = \operatorname*{argmax}_{\varkappa} L_1(\varkappa, \varkappa^*).$$

Problem: to bound excess loss

$$L_0(\widetilde{\varkappa}, \varkappa^*) \stackrel{\text{def}}{=} \frac{1}{2}\bigl\{\rho(\widetilde{\varkappa}) - \rho(\varkappa^*)\bigr\}.$$

## Rate function

For $L_1(\varkappa, \varkappa^*) \stackrel{\text{def}}{=} \frac{1}{2}\{\ell(\varkappa^*) - \ell(\varkappa)\}$ and a fixed $\mu_1 > 0$, define

$$\mathcal{M}_1(\mu_1, \varkappa, \varkappa^*) \stackrel{\text{def}}{=} -\log I\!\!E \exp\{\mu_1 L_1(\varkappa, \varkappa^*)\}$$

implying

$$I\!\!E \exp\Big\{\mu_1 L_1(\varkappa, \varkappa^*) + \mathcal{M}_1(\mu_1, \varkappa, \varkappa^*)\Big\} = 1.$$

Define the rate function

$$\mathcal{M}_1(\varkappa, \varkappa^*) = \mathcal{M}_1(\mu_1(\varkappa, \varkappa^*), \varkappa, \varkappa^*),$$

where $\mu_1(\varkappa, \varkappa^*)$ (nearly) maximizes $\mathcal{M}_1(\mu_1, \varkappa, \varkappa^*)$ w.r.t. $\mu_1$.

## Concentration of the empirical risk $\ell(\varkappa, \varkappa^*)$

### Theorem

*Let $\mathcal{K}$ be a discrete set and let $z(\varkappa, \varkappa^*)$ satisfy*

$$\sum_{\varkappa \in \mathcal{K}} \exp\{-z(\varkappa, \varkappa^*)\} = \mathfrak{Q}^* < \infty.$$

*Then*

$$\mathcal{C}_1(\varkappa, \varkappa^*) \stackrel{\text{def}}{=} \mathcal{M}_1(\varkappa, \varkappa^*) - z(\varkappa, \varkappa^*)$$

*is a concentration function for $L_1(\varkappa, \varkappa^*)$ in the sense that*

$$\boxed{\mathbb{E} \sup_{\varkappa \in \mathcal{K}} \exp\{\mu_1(\varkappa, \varkappa^*) L_1(\varkappa, \varkappa^*) + \mathcal{C}_1(\varkappa, \varkappa^*)\} \le \mathfrak{Q}^*.}$$

## Concentration sets

Concentration bound: for $\mathcal{C}_1(\varkappa, \varkappa^*) = \mathcal{M}_1(\varkappa, \varkappa^*) - z(\varkappa, \varkappa^*)$

$$\boxed{\mathbb{E} \sup_{\varkappa \in \mathcal{K}} \exp\bigl\{\mu_1(\varkappa, \varkappa^*) L_1(\varkappa, \varkappa^*) + \mathcal{C}_1(\varkappa, \varkappa^*)\bigr\} \le \mathfrak{Q}^*,}$$

Concentration set: $\mathcal{A}_1(\mathfrak{z}) \stackrel{\mathrm{def}}{=} \{\varkappa : \mathcal{C}_1(\varkappa, \varkappa^*) \le \mathfrak{z}\}$. By $L_1(\tilde{\varkappa}, \varkappa^*) \ge 0$:

$$\boxed{\mathbb{P}\bigl(\tilde{\varkappa} \notin \mathcal{A}_1(\mathfrak{z})\bigr) = \mathbb{P}\bigl(\mathcal{C}_1(\varkappa, \varkappa^*) > \mathfrak{z}\bigr) \le \mathfrak{Q}^* e^{-\mathfrak{z}}.}$$

Concentration of emp. risk: with $\mathfrak{b}_1(\mathfrak{z}) \stackrel{\mathrm{def}}{=} \min_{\varkappa \in \mathcal{K}} \bigl\{\mathfrak{z}\mu_1(\varkappa, \varkappa^*) + \mathcal{C}_1(\varkappa, \varkappa^*)\bigr\}$

$$\boxed{\mathbb{P}\bigl(L_1(\tilde{\varkappa}, \varkappa^*) > \mathfrak{z}\bigr) \le \mathfrak{Q}^* \exp\{-\mathfrak{b}_1(\mathfrak{z})\}.}$$

## Concentration of the loss

Consider

$$L_0(\varkappa, \varkappa^*) \stackrel{\text{def}}{=} \{\rho(\varkappa) - \rho(\varkappa^*)\}/2$$

$$\mathcal{M}_0(\mu_0, \varkappa, \varkappa^*) \stackrel{\text{def}}{=} -\log \mathbb{E} \exp\bigl\{\mu_0 L_0(\varkappa, \varkappa^*)\bigr\}.$$

Assume $\mu_0(\varkappa, \varkappa^*)$ is fixed for $\varkappa \in \mathcal{K}$ defining

$$\mathcal{M}_0(\varkappa, \varkappa^*) \stackrel{\text{def}}{=} \mathcal{M}_0(\mu_0(\varkappa, \varkappa^*), \varkappa, \varkappa^*) \ge 0.$$

For $\mathcal{C}_0(\varkappa, \varkappa^*) = \mathcal{M}_0(\varkappa, \varkappa^*) - z(\varkappa, \varkappa^*)$, it holds:

$$\boxed{\mathbb{E} \sup_{\varkappa \in \mathcal{K}} \exp\bigl\{\mu_0(\varkappa, \varkappa^*) L_0(\varkappa, \varkappa^*) + \mathcal{C}_0(\varkappa, \varkappa^*)\bigr\} \le \mathfrak{Q}^*.}$$

Yields concentration sets and concentration bound.

## Cross Concentration

Now consider

$$L_0(\varkappa, \varkappa^*) \stackrel{\text{def}}{=} \{\rho(\varkappa) - \rho(\varkappa^*)\}/2.$$

Aim: a bound for $L_0(\tilde{\varkappa}, \varkappa^*)$ with

$$\tilde{\varkappa} = \operatorname*{argmin}_{\varkappa} \ell(\varkappa) = \operatorname*{argmax}_{\varkappa} L_1(\varkappa, \varkappa^*).$$

Idea: consider linear combinations

$$\mathcal{L}(\varkappa, \varkappa^*) \stackrel{\text{def}}{=} \mu_1(\varkappa, \varkappa^*) L_1(\varkappa, \varkappa^*) + \mu_0(\varkappa, \varkappa^*) L_0(\varkappa, \varkappa^*).$$

Define the rate function

$$\mathcal{M}(\varkappa, \varkappa^*) \stackrel{\text{def}}{=} -\log \mathbb{E} \exp\bigl\{\mathcal{L}(\varkappa, \varkappa^*)\bigr\}.$$

## Concentration bound

### Theorem

*Let*

$$\sum_{\mathcal{K}} \exp\{-z(\varkappa, \varkappa^*)\} = \mathfrak{Q}^* < \infty.$$

*Then*

$$\mathcal{C}(\varkappa, \varkappa^*) \stackrel{\text{def}}{=} \mathcal{M}(\varkappa, \varkappa^*) - z(\varkappa, \varkappa^*)$$

*is a concentration function for* $\mathcal{L}(\varkappa, \varkappa^*)$ *in the sense that*

$$\boxed{\mathbb{E} \sup_{\varkappa \in \mathcal{K}} \exp\{\mathcal{L}(\varkappa, \varkappa^*) + \mathcal{C}(\varkappa, \varkappa^*)\} \leq \mathfrak{Q}^*.}$$

## Concentration of the excess loss

### Corollary

*Let*

$$\mathbb{E} \sup_{\varkappa \in \mathcal{K}} \exp\Bigl\{ \mathcal{L}(\varkappa, \varkappa^*) + \mathcal{C}(\varkappa, \varkappa^*) \Bigr\} \le \mathfrak{Q}^*.$$

*Define for* $\mathfrak{z} \ge 0$

$$\mathfrak{b}(\mathfrak{z}) \stackrel{\text{def}}{=} \inf_{\varkappa \in \mathcal{K}} \bigl\{ \mathfrak{z}\, \mu_0(\varkappa, \varkappa^*) + \mathcal{C}(\varkappa, \varkappa^*) \bigr\}.$$

*Then for any* $\mathfrak{z} \ge 0$ *, it holds for* $L_0(\widetilde{\varkappa}, \varkappa^*) = \{\rho(\widetilde{\varkappa}) - \rho(\varkappa^*)\}/2$

$$\boxed{\mathbb{P}\Bigl\{ L_0(\widetilde{\varkappa}, \varkappa^*) > \mathfrak{z} \Bigr\} = \mathbb{P}\Bigl\{ \rho(\widetilde{\varkappa}) - \rho(\varkappa^*) > 2\mathfrak{z} \Bigr\} \le \mathfrak{Q}^* e^{-\mathfrak{b}(\mathfrak{z})}.}$$

## Proof

From definition, $L_1(\widetilde{\varkappa}, \varkappa^*) \geq 0$ and

$$\mathfrak{b}(\mathfrak{z}) \leq \bigl\{\mathfrak{z}\, \mu_0(\widetilde{\varkappa}, \varkappa^*) + \mathcal{C}(\widetilde{\varkappa}, \varkappa^*)\bigr\}.$$

Therefore

$$\begin{aligned}
&I\!\!P\bigl\{L_0(\widetilde{\varkappa}, \varkappa^*) > \mathfrak{z}\bigr\} \\
&\qquad \leq I\!\!P\bigl\{\mu_0(\widetilde{\varkappa}, \varkappa^*)L_0(\widetilde{\varkappa}, \varkappa^*) + \mu_1(\widetilde{\varkappa}, \varkappa^*)L_1(\widetilde{\varkappa}, \varkappa^*) > \mu_0(\widetilde{\varkappa}, \varkappa^*)\mathfrak{z}\bigr\} \\
&\qquad \leq I\!\!P\bigl\{\mathcal{L}(\widetilde{\varkappa}, \varkappa^*) + \mathcal{C}(\widetilde{\varkappa}, \varkappa^*) > \mu_0(\widetilde{\varkappa}, \varkappa^*)\mathfrak{z} + \mathcal{C}(\widetilde{\varkappa}, \varkappa^*)\bigr\} \\
&\qquad \leq I\!\!P\bigl\{\mathcal{L}(\widetilde{\varkappa}, \varkappa^*) + \mathcal{C}(\widetilde{\varkappa}, \varkappa^*) > \mathfrak{b}(\mathfrak{z})\bigr\} \\
&\qquad \leq \mathrm{e}^{-\mathfrak{b}(\mathfrak{z})} I\!\!E \exp\bigl\{\mathcal{L}(\widetilde{\varkappa}, \varkappa^*) + \mathcal{C}(\widetilde{\varkappa}, \varkappa^*)\bigr\}.
\end{aligned}$$

## Projection estimates

Consider $\widetilde{s}_i(\varkappa) = \varkappa_i Y_i$, where $\varkappa = (\varkappa_i)$, with $\varkappa_i \in \{0,1\}$. Then

$$\rho(\varkappa) \; = \; \sum_i \bigl(\varkappa_i Y_i - s_i\bigr)^2 = \sum_{i:\, \varkappa_i = 0} s_i^2 + \sum_{i:\, \varkappa_i = 1} \bigl(Y_i - s_i\bigr)^2,$$

$$\mathcal{R}(\varkappa) \; = \; \mathbb{E} \sum_i \bigl(\varkappa_i Y_i - s_i\bigr)^2 = \sum_{i:\, \varkappa_i = 0} s_i^2 + \sum_{i:\, \varkappa_i = 1} \sigma_i^2$$

$$\mathcal{R}(\varkappa, \varkappa^*) \stackrel{\text{def}}{=} \mathcal{R}(\varkappa) - \mathcal{R}(\varkappa^*) = - \sum_i \delta_i (s_i^2 - \sigma_i^2) \ge 0,$$

where $\delta_i = \varkappa_i - \varkappa_i^* \in \{1, 0, -1\}$.

## One-sided concentration: $\varkappa > \varkappa^*$

Let $\varkappa^* = \operatorname{argmin}_{\varkappa} \mathcal{R}(\varkappa) = \operatorname{argmin}_{\varkappa} \mathcal{R}(\varkappa, \varkappa^*)$ .

Let $\varkappa \geq \varkappa'$ mean partial ordering. Consider

$$\mathcal{K}^+(\varkappa^*) = \{\varkappa \in \mathcal{K} : \varkappa > \varkappa^*\}.$$

Requirement to the procedure: restricted to $\mathcal{K}^+(\varkappa^*)$ , the selected model should be nearly oracle. Equivalently: the event

$$\{L_1(\varkappa, \varkappa^*) < 0\} = \{\ell(\varkappa) - \ell(\varkappa^*) < 0\}, \qquad \varkappa > \varkappa^*$$

only possible for $\varkappa \approx \varkappa^*$ . Expected if

$$\mathbb{E} L_1(\varkappa, \varkappa^*) = \mathcal{R}(\varkappa, \varkappa^*) \gg \operatorname{Var}^{1/2}\bigl(L_1(\varkappa, \varkappa^*)\bigr).$$

## $\varkappa > \varkappa^*$: loss and risk

We write $\varkappa \geq \varkappa^*$ if $\varkappa_i \geq \varkappa_i^*$ for all $i$. For projection estimates, this means $\mathcal{I}_{\varkappa^*} \subset \mathcal{I}_{\varkappa}$, where $\mathcal{I}_{\varkappa} = \{i : \varkappa_i = 1\}$.

Define

$$V^2(\varkappa, \varkappa^*) \stackrel{\text{def}}{=} \sum_{\mathcal{I}_{\varkappa} \setminus \mathcal{I}_{\varkappa^*}} \sigma_i^2, \qquad B(\varkappa, \varkappa^*) \stackrel{\text{def}}{=} \sum_{\mathcal{I}_{\varkappa} \setminus \mathcal{I}_{\varkappa^*}} s_i^2.$$

It holds

$$\mathcal{R}(\varkappa, \varkappa^*) = \sum_{\mathcal{I}_{\varkappa} \setminus \mathcal{I}_{\varkappa^*}} (-s_i^2 + \sigma_i^2) = V^2(\varkappa, \varkappa^*) - B(\varkappa, \varkappa^*).$$

The definition of $\varkappa^*$ implies $B(\varkappa, \varkappa^*) \leq V^2(\varkappa, \varkappa^*)$ and $\mathcal{R}(\varkappa, \varkappa^*) \leq V^2(\varkappa, \varkappa^*)$.

## $\varkappa > \varkappa^*$ : rate function

Fix $\mu_0, \mu_1$ and $\mathcal{L}(\varkappa, \varkappa^*) = \mu_0 L_0(\varkappa, \varkappa^*) + \mu_1 L_1(\varkappa, \varkappa^*)$ :

If $\mu_1 + \mu_0 < \sigma^{-2}(\varkappa, \varkappa^*)$ with $\sigma(\varkappa, \varkappa^*) = \max_{i \in \mathcal{I}_\varkappa} \sigma_i$ , then

$$\begin{aligned} I\!\!E \mathcal{L}(\mu_0, \mu_1, \varkappa, \varkappa^*) &= \frac{\mu_1 - \mu_0}{2} \mathcal{R}(\varkappa, \varkappa^*), \\ \operatorname{Var}\bigl(\mathcal{L}(\mu_0, \mu_1, \varkappa, \varkappa^*)\bigr) &\asymp (\mu_0 + \mu_1)^2 \sigma^2(\varkappa, \varkappa^*) V^2(\varkappa, \varkappa^*), \end{aligned}$$

and

$$\mathcal{M}(\mu_0, \mu_1, \varkappa, \varkappa^*) \geq \frac{\mu_1 - \mu_0}{2} \mathcal{R}(\varkappa, \varkappa^*) - C(\mu_0 + \mu_1)^2 \sigma^2(\varkappa, \varkappa^*) V^2(\varkappa, \varkappa^*).$$

## $\varkappa > \varkappa^*$ : correction for multiplicity

Let

$$\sum_{\varkappa > \varkappa^*} \exp\bigl\{-z(\varkappa, \varkappa^*)\bigr\} \le \mathfrak{Q}^+.$$

Define

$$\mu(\varkappa, \varkappa^*) \stackrel{\text{def}}{=} \frac{\sqrt{z(\varkappa, \varkappa^*)}}{\sigma(\varkappa, \varkappa^*) V(\varkappa, \varkappa^*)} \qquad \varkappa > \varkappa^*,$$

and for some $\alpha_0 \le \alpha_1$ , set

$$\mu_0(\varkappa, \varkappa^*) = \alpha_0 \mu(\varkappa, \varkappa^*) \qquad \mu_1(\varkappa, \varkappa^*) = \alpha_1 \mu(\varkappa, \varkappa^*)$$

Denote

$$\mathcal{M}(\varkappa, \varkappa^*; \alpha_0, \alpha_1) \stackrel{\text{def}}{=} \mathcal{M}\bigl(\alpha_0 \mu(\varkappa, \varkappa^*), \alpha_1 \mu(\varkappa, \varkappa^*), \varkappa, \varkappa^*\bigr).$$

## $\varkappa > \varkappa^*$ : One-sided concentration

### Theorem

*Let* $\sum_{\varkappa > \varkappa^*} \exp\{-z(\varkappa, \varkappa^*)\} \le \mathfrak{Q}^+$ . *Let* $\alpha_1 > \alpha_0$ *be s.t.*

$$(\alpha_0 + \alpha_1)\mu(\varkappa, \varkappa^*)\sigma^2(\varkappa, \varkappa^*) = \frac{(\alpha_0 + \alpha_1)\sigma(\varkappa, \varkappa^*)\sqrt{z(\varkappa, \varkappa^*)}}{V(\varkappa, \varkappa^*)} \le c < 1,$$

*Then*

$$I\!P\Bigl(\rho(\widetilde{\varkappa}) - \rho(\varkappa^*) \ge 2\mathfrak{z}, \widetilde{\varkappa} > \varkappa^*\Bigr) \le \mathfrak{Q}^+ \exp\{-\mathfrak{b}^+(\mathfrak{z}; \alpha_0, \alpha_1)\}$$

*with*

$$\mathfrak{b}^+(\mathfrak{z}; \alpha_0, \alpha_1) \stackrel{\text{def}}{=} \min_{\varkappa > \varkappa^*} \bigl\{\mathfrak{z}\,\alpha_0\,\mu(\varkappa, \varkappa^*) + \mathcal{M}(\varkappa, \varkappa^*; \alpha_0, \alpha_1) - z(\varkappa, \varkappa^*)\bigr\}.$$

## $\varkappa > \varkappa^*$ : Conditions on the risk function

### Theorem

*If for every* $\varkappa > \varkappa^*$

$$\boxed{\mathcal{R}(\varkappa, \varkappa^*) \geq \mathfrak{t}\, \sigma(\varkappa, \varkappa^*)\, V(\varkappa, \varkappa^*) \sqrt{z(\varkappa, \varkappa^*)},}$$

*where the constant* $\mathfrak{t}$ *is sufficiently large to ensure that*

$$\delta \stackrel{\text{def}}{=} \frac{1}{2}\big[\mathfrak{t}(\alpha_1 - \alpha_0) - q^+(c)(\alpha_1 + \alpha_0)^2 - (1-c)^{-1}\alpha_1^2\big] - 1 > 0,$$

*then*

$$\mathfrak{b}^+(\mathfrak{z}; \alpha_0, \alpha_1) \geq \min_{\varkappa > \varkappa^*} \left\{ \frac{\mathfrak{z}\alpha_0 \sqrt{z(\varkappa, \varkappa^*)}}{\sigma(\varkappa, \varkappa^*) V(\varkappa, \varkappa^*)} + \delta z(\varkappa, \varkappa^*) \right\}.$$

## Discussion

SURE/AIC works if $\mathcal{R}(\varkappa, \varkappa^*)$ grows sufficiently fast, so that

$$\mathcal{R}(\varkappa, \varkappa^*) \geq \mathfrak{t}\, \sigma(\varkappa, \varkappa^*)\, V(\varkappa, \varkappa^*) \sqrt{z(\varkappa, \varkappa^*)}.$$

SURE/AIC does not work if

- $\mathcal{R}(\varkappa)$ is flat and thus $\mathcal{R}(\varkappa, \varkappa^*)$ is small.
- $\sigma_{\varkappa}^2 \asymp V^2(\varkappa, \varkappa^*)$ i.e. for severely ill-posed cases.

Interpretation: Risk is significant relative to the variation of the empirical risk: $\operatorname{Var}(L_1(\varkappa, \varkappa^*)) \asymp \sigma^2(\varkappa, \varkappa^*)\, V^2(\varkappa, \varkappa^*)$ .

## $\varkappa < \varkappa^*$: loss and risk

$\varkappa < \varkappa^*$ means $\mathcal{I}_{\varkappa} \subset \mathcal{I}_{\varkappa^*}$ .

The excess risk $\mathcal{R}(\varkappa, \varkappa^*)$ can be represented as

$$\mathcal{R}(\varkappa, \varkappa^*) = -B(\varkappa, \varkappa^*) - V^2(\varkappa, \varkappa^*)$$

where $V^2(\varkappa, \varkappa^*) = V^2(\varkappa^*, \varkappa) = V^2(\varkappa^*) - V^2(\varkappa)$ and $B(\varkappa, \varkappa^*) = B(\varkappa^*) - B(\varkappa)$ .

The definition of $\varkappa^*$ implies that $-B(\varkappa, \varkappa^*) > V^2(\varkappa, \varkappa^*)$ .

## $\varkappa < \varkappa^*$: rate functions

With $\mathcal{I}(\varkappa^*, \varkappa) = \mathcal{I}(\varkappa^*) \setminus \mathcal{I}(\varkappa)$, it holds

$$\mathcal{M}_0(\mu_0, \varkappa, \varkappa^*) = -\frac{\mu_0}{2}\mathcal{R}(\varkappa, \varkappa^*) - \frac{1}{2} \sum_{\mathcal{I}(\varkappa^*, \varkappa)} Q(-\mu_1 \sigma_i^2),$$

$$\mathcal{M}_1(\mu_1, \varkappa, \varkappa^*) = \frac{\mu_1}{2}\mathcal{R}(\varkappa, \varkappa^*) - \frac{1}{2} \sum_{\mathcal{I}(\varkappa^*, \varkappa)} \left\{ Q(-\mu_1 \sigma_i^2) + \frac{\mu_1^2 s_i^2 \sigma_i^2}{1 + \mu_1 \sigma_i^2} \right\}$$

$$\mathcal{M}(\mu_0, \mu_1, \varkappa, \varkappa^*) = \frac{\mu_1 - \mu_0}{2}\mathcal{R}(\varkappa, \varkappa^*) - \frac{1}{2} \sum_{\mathcal{I}(\varkappa^*, \varkappa)} \left\{ Q\bigl(-(\mu_1 + \mu_0)\sigma_i^2\bigr) + \frac{\mu_1^2 s_i^2 \sigma_i^2}{1 + (\mu_1 + \mu_0)\sigma_i^2} \right\}.$$

## $\varkappa < \varkappa^*$

Define $z^*(\varkappa^*) = \max_{\varkappa < \varkappa^*} z(\varkappa, \varkappa^*)$ and set for $\varkappa < \varkappa^*$

$$\mu(\varkappa, \varkappa^*) \equiv \mu^-(\varkappa^*) \stackrel{\text{def}}{=} \frac{\sqrt{z^*(\varkappa^*)}}{\sigma(\varkappa^*) V(\varkappa^*)}$$

with $\sigma(\varkappa^*) \stackrel{\text{def}}{=} \max_{\mathcal{I}(\varkappa^*)} \sigma_i$ . Also

$$\mu_0(\varkappa, \varkappa^*) = \alpha_0 \mu(\varkappa, \varkappa^*) \qquad \mu_1(\varkappa, \varkappa^*) = \alpha_1 \mu(\varkappa, \varkappa^*)$$

for some $\alpha_0 \le \alpha_1$ . Write

$$\mathcal{M}(\varkappa, \varkappa^*; \alpha_0, \alpha_1) \stackrel{\text{def}}{=} \mathcal{M}\bigl(\alpha_0 \mu(\varkappa, \varkappa^*), \alpha_1 \mu(\varkappa, \varkappa^*), \varkappa, \varkappa^*\bigr).$$

## Theorem ( $\varkappa < \varkappa^*$ : One-sided concentration)

*Let* $\sum_{\varkappa < \varkappa^*} \exp\bigl\{-z(\varkappa, \varkappa^*)\bigr\} \le \mathfrak{Q}^-$ , $z^*(\varkappa^*) = \max_{\varkappa < \varkappa^*} z(\varkappa, \varkappa^*)$ , *and* $(\alpha_0 + \alpha_1)\sqrt{z^*(\varkappa^*)}\sigma(\varkappa^*)/V(\varkappa^*) \le c < 1$ . *Define*

$$\mathfrak{b}^-(\mathfrak{z}; \alpha_0, \alpha_1) \stackrel{\text{def}}{=} \min_{\varkappa < \varkappa^*} \bigl\{\mathfrak{z}\, \mu_0(\varkappa, \varkappa^*) + \mathcal{M}(\varkappa, \varkappa^*; \alpha_0, \alpha_1) - z(\varkappa, \varkappa^*)\bigr\}.$$

*Then for any* $u > 0$

$$\begin{aligned} I\!P\Bigl(\frac{1}{\sigma(\varkappa^*)V(\varkappa^*)\sqrt{z^*(\varkappa^*)}}\bigl[\rho(\widetilde{\varkappa}) - \rho(\varkappa^*)\bigr] \ge 2u,\ \widetilde{\varkappa} < \varkappa^*\Bigr) \\ \le \mathfrak{Q}^- \exp\bigl\{\mathfrak{b}^-(\mathfrak{z}; \alpha_0, \alpha_1)\bigr\} \\ \le \mathfrak{Q}^- \exp\Bigl\{-\bigl[u\alpha_0 - g^-(c)\,(\alpha_0 + \alpha_1)^2/2 - 1\bigr] z^*(\varkappa^*)\Bigr\}. \end{aligned}$$

## Spectral cut-off procedure

$\varkappa$ is an integer and $\mathcal{I}_{\varkappa} = \{1, \ldots, \varkappa\}$, that is,

$$\widetilde{s}_i(\varkappa) = \begin{cases} Y_i & i \le \varkappa, \\ 0 & \text{otherwise.} \end{cases}$$

For a specific $\varkappa$, the corresponding loss and risk read as

$$\rho(\varkappa) = \sum_{i \le \varkappa} \bigl(Y_i - s_i\bigr)^2 + \sum_{i > \varkappa} s_i^2 \qquad \mathcal{R}(\varkappa) = \sum_{i \le \varkappa} \sigma_i^2 + \sum_{i > \varkappa} s_i^2 .$$

For the empirical risk holds

$$\ell(\varkappa) = \sum_{i \le \varkappa} \sigma_i^2 + \sum_{i > \varkappa} (Y_i^2 - \sigma_i^2), \qquad \mathbb{E}\ell(\varkappa) = \mathcal{R}(\varkappa).$$

Optimal choice: $$\varkappa^* = \operatorname*{argmin}_{\varkappa \in \mathcal{K}} \mathcal{R}(\varkappa) = \operatorname*{argmin}_{\varkappa \in \mathcal{K}} \Bigl\{ \sum_{i < \varkappa} \sigma_i^2 + \sum_{i > \varkappa} s_i^2 \Bigr\}$$

## Oracle two-sided bound

Let $z_0(k)$ satisfy

$$\sum_{k=1}^{\infty} \exp\{-z_0(k)\} = \mathfrak{Q}.$$

An example: $z_0(k) = (1+\delta)\log(k)$ for $\delta > 0$. Define

$$z(\varkappa, \varkappa^*) = z_0\bigl(k(\varkappa, \varkappa^*)\bigr), \qquad k(\varkappa, \varkappa^*) = \#\bigl\{\mathcal{K}(\varkappa, \varkappa^*)\bigr\}. \tag{2}$$

### Theorem

*Let the set* $\mathcal{K}$ *be ordered and let* $z(\varkappa, \varkappa^*)$ *be defined by* (2). *Then*

$$I\!\!P\Bigl(\rho(\widetilde{\varkappa}) - \rho(\varkappa^*) \ge 2\mathfrak{z}\Bigr) \le \mathfrak{Q} e^{-\mathfrak{b}^+(\mathfrak{z})} + \mathfrak{Q} e^{-\mathfrak{b}^-(\mathfrak{z})}$$

## Problem of choosing a subset

$\mathcal{K}$ , a set of subsets in $\mathcal{I}$ .

$\varkappa < \varkappa'$ means $\mathcal{I}_{\varkappa} \subset \mathcal{I}_{\varkappa'}$ .

Let $\varkappa^* = \operatorname{argmin}_{\mathcal{I}} \mathcal{R}(\varkappa)$ be the oracle choice.

Idea of the study: reduce the general oracle bound to the cases with $\varkappa < \varkappa^*$ and $\varkappa > \varkappa^*$ .

Approach: given $\varkappa, \varkappa^*$ consider ordered pairs $(\varkappa \vee \varkappa^*, \varkappa^*)$ and $(\varkappa^*, \varkappa \wedge \varkappa^*)$ :

$$\varkappa \bigtriangleup \varkappa^* = \mathcal{I}(\varkappa \vee \varkappa^*, \varkappa^*) \cup \mathcal{I}(\varkappa^*, \varkappa^* \wedge \varkappa)$$

## Construction of $z(\varkappa, \varkappa^*)$

Suppose that two functions $z^+(\varkappa, \varkappa^*)$ for $\varkappa > \varkappa^*$ and $z^-(\varkappa, \varkappa^*)$ for $\varkappa < \varkappa^*$ be fixed in a way that

$$\sum_{\varkappa > \varkappa^*} \exp\bigl\{-z^+(\varkappa, \varkappa^*)\bigr\} \le \mathfrak{Q}^+, \quad \sum_{\varkappa < \varkappa^*} \exp\bigl\{-z^-(\varkappa, \varkappa^*)\bigr\} \le \mathfrak{Q}^-$$

for some fixed constants $\mathfrak{Q}^\pm$. Define for any $\varkappa \in \mathcal{K}$

$$z(\varkappa, \varkappa^*) \stackrel{\text{def}}{=} z^+(\varkappa \vee \varkappa^*, \varkappa^*) + z^-(\varkappa \wedge \varkappa^*, \varkappa^*).$$

Obviously

$$\sum_{\varkappa} \exp\bigl\{-z(\varkappa, \varkappa^*)\bigr\} \le \mathfrak{Q}^+\mathfrak{Q}^-.$$

## Theorem (Oracle bound)

*Let* $\alpha$ *be s.t.* $2\alpha\sqrt{z(\varkappa,\varkappa^*)}\sigma(\varkappa,\varkappa^*)/V(\varkappa,\varkappa^*) \le c < 1$ *for* $\varkappa < \varkappa^*$ *.*
*Then for any* $u > 0$ *and* $\mathfrak{z} = u\sigma(\varkappa^*)V(\varkappa^*)\sqrt{z^*(\varkappa^*)}$ *:*

$$
\begin{aligned}
I\!P\Bigl(&\frac{1}{\sigma(\varkappa^*)V(\varkappa^*)\sqrt{z^*(\varkappa^*)}}\bigl[\rho(\widetilde{\varkappa}) - \rho(\varkappa^*)\bigr] \ge 2u,\, \widetilde{\varkappa} < \varkappa^*\Bigr) \\
&\le \mathfrak{Q}^- \exp\bigl\{-\mathfrak{b}^-(\mathfrak{z};\alpha,\alpha)\bigr\} \\
&\le \mathfrak{Q}^- \exp\Bigl\{-\bigl[u\alpha - 2g^-(c)\,\alpha^2 - 1\bigr] z^*(\varkappa^*)\Bigr\}.
\end{aligned}
$$

*Let* $(\alpha_0 + \alpha_1)\sqrt{z(\varkappa,\varkappa^*)}\sigma(\varkappa,\varkappa^*)/V(\varkappa,\varkappa^*) < 1$ *for* $\varkappa > \varkappa^*$ *. Then*

$$
\begin{aligned}
I\!P\Bigl(&\rho(\widetilde{\varkappa}) - \rho(\varkappa^*) \ge 2\mathfrak{z}^- + 2\mathfrak{z}^+,\, \widetilde{\varkappa} \vee \varkappa^* > \varkappa^*\Bigr) \\
&\le \mathfrak{Q}^+\mathfrak{Q}^- \exp\bigl\{-\mathfrak{b}^+(\mathfrak{z}^+;\alpha_0,\alpha_1) - \mathfrak{b}^-(\mathfrak{z}^-;\alpha_0,\alpha_1)\bigr\}
\end{aligned}
$$

Define $\mu(\varkappa, \varkappa^*) = \sqrt{z(\varkappa, \varkappa^*)}\sigma(\varkappa, \varkappa^*)/V(\varkappa, \varkappa^*)$, $\varkappa > \varkappa^*$.

## Theorem (cont)

*Let the risk excess* $\mathcal{R}(\varkappa, \varkappa^*)$ *fulfill*

$$\mu(\varkappa, \varkappa^*)\mathcal{R}(\varkappa, \varkappa^*) \ge \mathfrak{t}\, z(\varkappa, \varkappa^*), \qquad \varkappa > \varkappa^*$$

*where the constant* $\mathfrak{t}$ *is sufficiently large to ensure*

$$\delta \stackrel{\mathrm{def}}{=} \frac{1}{2}\bigl[\mathfrak{t}(\alpha_1 - \alpha_0) - q^+(c)(\alpha_1 + \alpha_0)^2 - (1-c)^{-1}\alpha_1^2\bigr] - 1 > 0.$$

*Then*

$$\mathfrak{b}^+(\mathfrak{z}; \alpha_0, \alpha_1) \ge \min_{\varkappa > \varkappa^*} \bigl\{\mathfrak{z}\alpha_0\mu(\varkappa, \varkappa^*) + \delta z(\varkappa, \varkappa^*)\bigr\}.$$

## Outline

## SPM: discussion

New procedure should overcome the main drawbacks of the SURE/AIC approach.

Ultimate Goal:

- establish a concentration property (oracle bound) regardless degree of ill-posedness and risk flatness.
- suggest a unified principle for choosing a penalty function.

## Risk minimization vs saddle point

Consider the risk difference as functions of two arguments $\varkappa, \varkappa'$ :

$$\mathcal{R}(\varkappa, \varkappa') \stackrel{\text{def}}{=} \mathcal{R}(\varkappa) - \mathcal{R}(\varkappa').$$

The definition of $\varkappa^*$ as the minimizer of the risk function $\mathcal{R}(\varkappa)$ can be rewritten in the form:

$$\varkappa^* = \operatorname*{argmin}_{\varkappa} \mathcal{R}(\varkappa, \varkappa^*) = \operatorname*{argmax}_{\varkappa} \mathcal{R}(\varkappa^*, \varkappa).$$

In other words, $(\varkappa^*, \varkappa^*)$ is a saddle point of the function $\mathcal{R}(\varkappa, \varkappa')$ :

$$0 \equiv \mathcal{R}(\varkappa^*, \varkappa^*) \geq \mathcal{R}(\varkappa^*, \varkappa), \quad 0 \equiv \mathcal{R}(\varkappa^*, \varkappa^*) \leq \mathcal{R}(\varkappa, \varkappa^*), \quad \forall \varkappa.$$

## Model selection as a SP-problem

Basic observation: the minimizer of SURE $\ell(\varkappa)$ is a saddle point of $\ell(\varkappa, \varkappa') = \ell(\varkappa) - \ell(\varkappa')$.

SP-Approach: consider a penalized bivariate function

$$\ell(\varkappa, \varkappa') + 2\,\mathrm{PEN}(\varkappa, \varkappa') = \ell(\varkappa) - \ell(\varkappa') + 2\,\mathrm{PEN}(\varkappa, \varkappa')$$

and define the penalized choice of $\widehat{\varkappa}$ via the smallest saddle point:

$$\widehat{\varkappa} \stackrel{\mathrm{def}}{=} \operatorname*{argmin}_{\varkappa} \max_{\varkappa'} \{\ell(\varkappa, \varkappa') - 2\,\mathrm{PEN}(\varkappa, \varkappa')\}.$$

In other words, the value $\widehat{\varkappa}$ is a SP if it holds for any $\varkappa^{\circ}$:

$$\max_{\varkappa} \{\ell(\widehat{\varkappa}, \varkappa) - 2\,\mathrm{PEN}(\widehat{\varkappa}, \varkappa)\} \le \max_{\varkappa} \{\ell(\varkappa^{\circ}, \varkappa) - 2\,\mathrm{PEN}(\varkappa^{\circ}, \varkappa)\}.$$

## The choice of a bivariate penalty function $\mathrm{PEN}(\varkappa, \varkappa')$

Main requirements:

*concentration property on each set* $\varkappa \geq \varkappa^*$ .

Is granted if a uniform exponential bound can be established for the linear combination

$$\mathcal{L}(\varkappa, \varkappa^*) = \mu_1(\varkappa, \varkappa^*) L_1(\varkappa, \varkappa^*) + \mu_0(\varkappa, \varkappa^*) L_0(\varkappa, \varkappa^*)$$

e.g. in the form

$$I\!\!E \sup_{\varkappa > \varkappa^*} \exp\bigl\{\mathcal{L}(\varkappa, \varkappa^*) - \mathrm{PEN}(\varkappa, \varkappa^*)\bigr\} \leq \mathfrak{Q}^*.$$

## Construction of Penalty: ordered case

Suppose that for every $\varkappa^* \in \mathcal{K}$

$$\sum_{\mathcal{K}} \exp\bigl\{-z(\varkappa, \varkappa^*)\bigr\} \le \mathfrak{Q}^* < \infty, \qquad \forall \varkappa^* \in \mathcal{K}.$$

Define a penalty function for $\varkappa > \varkappa'$ :

$$\mathrm{PEN}(\varkappa, \varkappa') \stackrel{\text{def}}{=} \tau \sqrt{z(\varkappa, \varkappa')}\sigma(\varkappa, \varkappa')V(\varkappa, \varkappa') \qquad \varkappa > \varkappa',$$

## Construction of Penalty: unordered case

For any $\varkappa, \varkappa'$, the value $\mathrm{PEN}(\varkappa, \varkappa')$ is defined as:

$$\mathrm{PEN}(\varkappa, \varkappa') \stackrel{\text{def}}{=} \mathrm{PEN}(\varkappa \vee \varkappa', \varkappa')$$

Interpretation: $\mathrm{PEN}(\varkappa, \varkappa')$ reflects increased complexity of $\varkappa$ relative to $\varkappa'$.

- $\mathrm{PEN}(\varkappa \vee \varkappa', \varkappa')$ accounts for the increase of complexity by adding new components $\varkappa \setminus \varkappa'$ to $\varkappa'$.

Define

$$
\begin{aligned}
\mu(\varkappa, \varkappa') &= \frac{\sqrt{z(\varkappa, \varkappa')}}{\sigma(\varkappa, \varkappa') V(\varkappa, \varkappa')}\,, \\
\mu_0(\varkappa, \varkappa^*) &= \alpha_0 \mu(\varkappa, \varkappa^*), \\
\mu_1(\varkappa, \varkappa^*) &= \alpha_1 \mu(\varkappa, \varkappa^*),
\end{aligned}
$$

where $\alpha_0 \le \alpha_1$ are some constants. Set

$$
\begin{aligned}
\mathcal{L}(\varkappa, \varkappa^*; \alpha_0, \alpha_1) &= \mu_1(\varkappa, \varkappa^*) L_1(\varkappa, \varkappa^*) + \mu_0(\varkappa, \varkappa^*) L_0(\varkappa, \varkappa^*) \\
\mathcal{M}(\varkappa, \varkappa^*; \alpha_0, \alpha_1) &= -\log I\!\!E \exp\bigl\{\mathcal{L}(\varkappa, \varkappa^*; \alpha_0, \alpha_1)\bigr\}
\end{aligned}
$$

### Theorem (Oracle concentration bound)

*Let* $(\alpha_0 + \alpha_1)\mu(\varkappa, \varkappa^*)\sigma^2(\varkappa, \varkappa^*) \le c < 1$ . *Then it holds*

$$
\begin{aligned}
&I\!\!P\Big(\rho(\widetilde{\varkappa}) - \rho(\varkappa^*) \ge 2\mathfrak{z}^- + 2\mathfrak{z}^+ + 2\mathfrak{z}_1\Big) \\
&\qquad \le \mathfrak{Q}^- \exp\big\{-\mathfrak{b}^-(\mathfrak{z}^-;\alpha_0,\alpha_1)\big\} \\
&\qquad\quad + \mathfrak{Q}^+\mathfrak{Q}^- \exp\big\{-\mathfrak{b}^+(\mathfrak{z}^+;\alpha_0,\alpha_1) - \mathfrak{b}^-(\mathfrak{z}^-;\alpha_0,\alpha_1)\big\} \\
&\qquad\quad + \mathfrak{Q}^+\mathfrak{Q}^- \exp\big\{-\mathfrak{b}_1(\mathfrak{z}_1;\alpha_1)\big\}.
\end{aligned}
$$

## Theorem (Oracle concentration bound. cont)

*Let also the constant* $\tau$ *in the definition*

$$\mathrm{PEN}(\varkappa, \varkappa^*) = \tau \sqrt{z(\varkappa, \varkappa')} \sigma(\varkappa, \varkappa') V(\varkappa, \varkappa'), \qquad \varkappa > \varkappa',$$

*fulfill*

$$\delta \stackrel{\text{def}}{=} \frac{1}{2} \bigl[ \tau \alpha_1 - q^+(c)(\alpha_1 + \alpha_0)^2 - (1-c)^{-1} \alpha_1^2 \bigr] - 1 > 0.$$

*Then*

$$\mathfrak{b}_1(\mathfrak{z}; \alpha_1) \ge \mathfrak{b}^+(\mathfrak{z}; \alpha_0, \alpha_1) \ge \min_{\varkappa > \varkappa^*} \bigl\{ \mathfrak{z} \alpha_0 \mu(\varkappa, \varkappa^*) + \delta z(\varkappa, \varkappa^*) \bigr\}.$$

## Ordered model choice: $\mathcal{K} = \{1, 2, \ldots, K\}$

Homogeneous noise: $\sigma_i = \sigma$ , $V^2(\varkappa, \varkappa^*) = \sigma^2 |\varkappa - \varkappa^*|$ . The choice

$$\mathrm{PEN}(\varkappa, \varkappa') = C \sqrt{\log(|\varkappa - \varkappa'|)} \qquad \varkappa > \varkappa'$$

yields the bound for excess loss of order $\sigma(\varkappa^*) V(\varkappa^*) \sqrt{\log n}$ .

Inverse problem: $\sigma_i$ grows with $i$ . Then $\sigma(\varkappa, \varkappa^*) = \max_{i \in \varkappa \vee \varkappa^*} \sigma_i$ and

$$\mathrm{PEN}(\varkappa, \varkappa^*) = C \sqrt{\sigma(\varkappa, \varkappa') V(\varkappa, \varkappa') \log(|\varkappa - \varkappa'|)} \qquad \varkappa > \varkappa'$$

still yields the oracle bound for $\sigma(\varkappa, \varkappa') \ll V(\varkappa, \varkappa')$ .

Severely ill-posed case: $\sigma(\varkappa, \varkappa^*) \asymp V(\varkappa, \varkappa^*)$ .
Then $L_0(\widetilde{\varkappa}, \varkappa^*) \asymp \sqrt{\sigma(\varkappa^*) V(\varkappa^*) \log(n)} \gg \mathcal{R}(\varkappa^*) \asymp V^2(\varkappa^*)$ .

## Subset selection

$\mathcal{K} =$ "set of all subsets" + "homogeneous noise" $\sigma_i \equiv \sigma$ .

The choice

$$\mathrm{PEN}(\varkappa, \varkappa') = C\sigma \sqrt{(|\varkappa| - |\varkappa'|) \log(n - |\varkappa'|)} \qquad \varkappa > \varkappa'$$

yields the oracle bound.

The excess loss bound is of order $|\varkappa^*| \sqrt{\log(n - |\varkappa^*|)}$ , much larger than the oracle risk $\mathcal{R}(\varkappa^*) \asymp |\varkappa^*|$ .

## Outlooks

- New SP-approach is only relevant for practical use, if an efficient implementation is available. A choice of a SP can be reduced to a semi-definite programming provided that the objective function is convex-concave. This seems to be possible to achieve for typical examples.

- Extensions to linear models, non-Gaussian/correlated errors, etc. are called for.

- Numerical examples still missing.