Sélection de modèles et sélection d'estimateurs pour l'Apprentissage statistique

Sylvain Arlot

[1]Cnrs

[2]École Normale Supérieure (Paris), Liens, Équipe Sierra

Cours Peccot, Collège de France, 17/01/2011

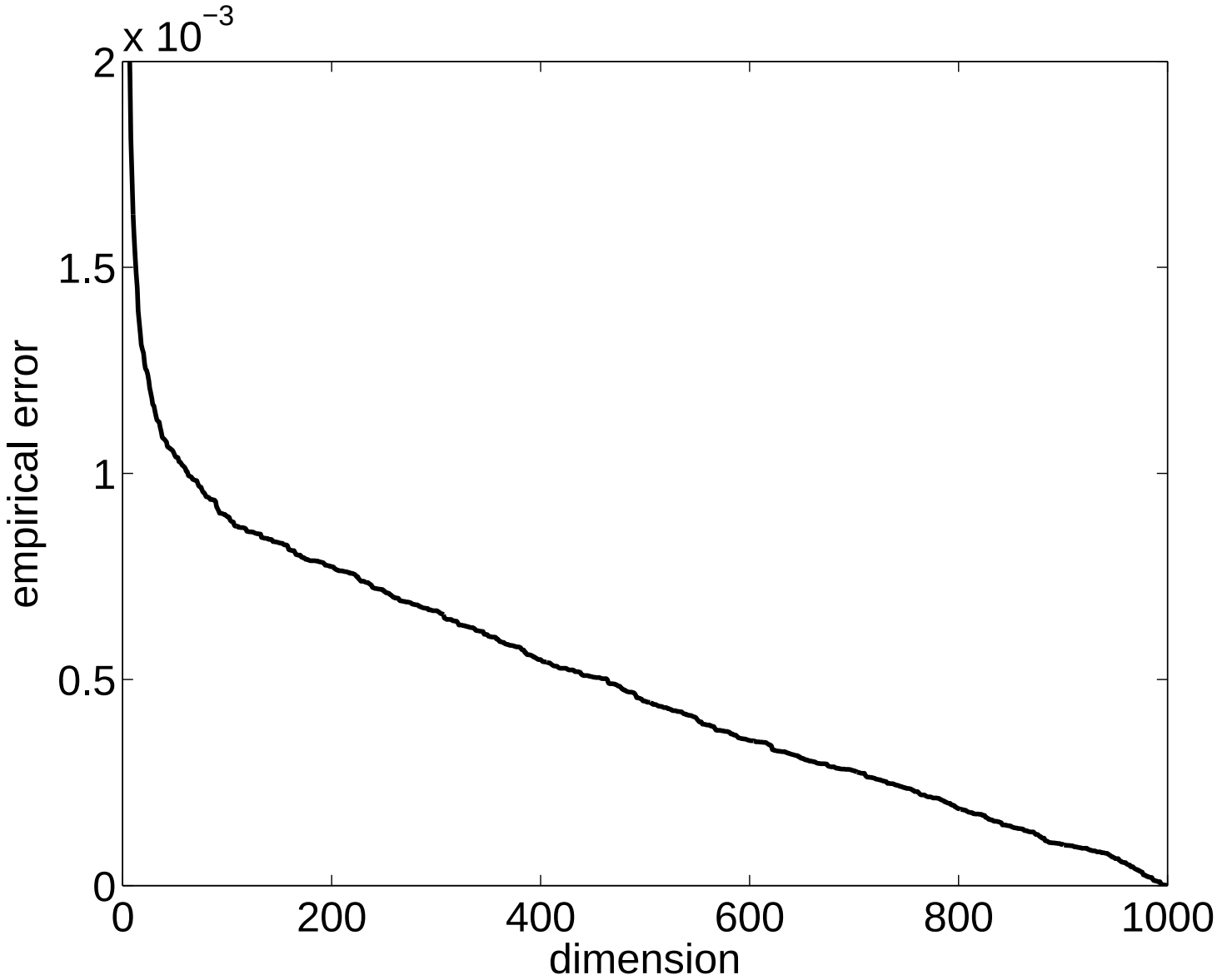
Plan

1. Lundi 10 : Apprentissage statistique et sélection d'estimateurs
2. Aujourd'hui : Calibration de pénalités et pénalités minimales
3. Lundi 24, 14h–16h : Rééchantillonnage et pénalisation
4. Lundi 31, 15h30–17h30 : Validation croisée et pénalités reliées

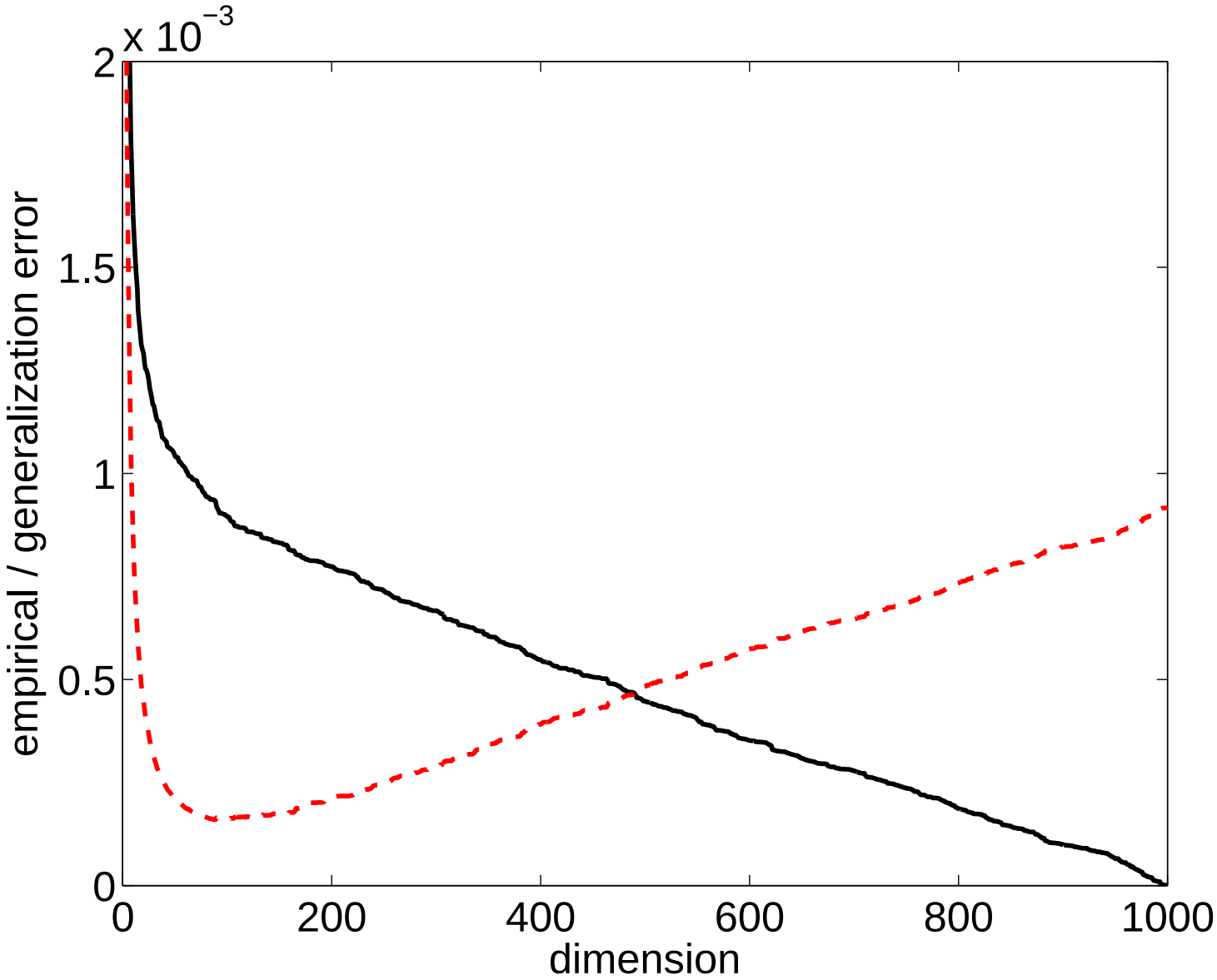
Plan du cours

Plan

Motivations (1) : "L-curve" et heuristique de coude ?



Motivations (1) : "L-curve" et heuristique de coude ?



Motivations (2) : que se passe-t-il si $K \leq 1$?

Théorème (Birgé & Massart 2007)

Si $K > 1$, pour tout $x \geq 0$, avec probabilité $1 - 4\,\mathrm{Card}(\mathcal{M}_n)e^{-x}$, pour tout

$$\widehat{m} \in \arg\min_{m \in \mathcal{M}_n} \left\{ \frac{1}{n} \left\| Y - \widehat{F}_m \right\|^2 + \frac{K\sigma^2 \dim(S_m)}{n} \right\} ,$$

on a l'inégalité-oracle $\forall \delta > 0$,

$$\frac{1}{n} \left\| \widehat{F}_{\widehat{m}} - F \right\|^2 \leq \left(\frac{1 + (K-2)_+}{1 - (2-K)_+} + \delta \right) \inf_{m \in \mathcal{M}_n} \left\{ \frac{1}{n} \left\| \widehat{F}_m - F \right\|^2 \right\} + \frac{C(K) x \sigma^2}{\delta n}$$

Motivations (3) : calibration de pénalités

- C_p et C_L (Mallows, 1973) :

$$\text{pen}(m) = \frac{2\sigma^2 D_m}{n}$$

$$\text{pen}(m) = \frac{2\sigma^2 \,\text{tr}(A_m)}{n}$$

- Pénalités proportionnelles à D_m avec constante multiplicative optimale inconnue : détection de ruptures (Birgé & Massart, 2001 ; Lebarbier, 2005), modèles de mélange (Maugis & Michel, 2008), etc.
- Pénalités de Rademacher

$$\text{pen}(m) = 2 \times \mathbb{E}\left[\sup_{t \in S_m} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\varepsilon_i \gamma\left(t; \xi_i \right) \right) \middle| \; D_n \right\} \right]$$

- ...

Estimateur naïf de σ^2

Exemple : régression homoscédastique sur un design fixe

Calcul du risque empirique

Estimateur naïf de σ^2

Exemple : régression homoscédastique sur un design fixe

$$\mathbb{E}\left[\frac{1}{n}\left\|Y-\widehat{F}_m\right\|^2\right]=\frac{1}{n}\left\|(I_n-A_m)F\right\|^2+\frac{\sigma^2(n-D_m)}{n}$$

Estimateur naïf de σ^2 :

$$\widehat{\sigma}^2_m:=\frac{1}{n-D_m}\left\|Y-\widehat{F}_m\right\|^2$$

Biais de cet estimateur :

$$\mathbb{E}\left[\widehat{\sigma}^2_m\right]=\sigma^2+\frac{1}{n-D_m}\left\|(I_n-A_m)F\right\|^2 \quad ,$$

Estimateur naïf de σ^2

Exemple : régression homoscédastique sur un design fixe

$$\mathbb{E}\left[\frac{1}{n}\left\|Y-\widehat{F}_m\right\|^2\right]=\frac{1}{n}\left\|(I_n-A_m)F\right\|^2+\frac{\sigma^2(n-D_m)}{n}$$

Estimateur naïf de σ^2 :

$$\widehat{\sigma}^2_m:=\frac{1}{n-D_m}\left\|Y-\widehat{F}_m\right\|^2$$

Biais de cet estimateur :

$$\mathbb{E}\left[\widehat{\sigma}^2_m\right]=\sigma^2+\frac{1}{n-D_m}\left\|(I_n-A_m)F\right\|^2 \quad ,$$

$\Rightarrow$ Utilisation dans la pénalité $2\sigma^2 D_m/n$?

Estimateur naïf de σ^2

Estimateur naïf de σ^2 :

$$\widehat{\sigma}_m^2 := \frac{1}{n - D_m} \left\| Y - \widehat{F}_m \right\|^2$$

Biais de cet estimateur :

$$\mathbb{E}\left[\widehat{\sigma}_m^2\right] = \sigma^2 + \frac{1}{n - D_m} \left\| (I_n - A_m) F \right\|^2 \quad ,$$

$\Rightarrow$ Utilisation dans la pénalité $2\sigma^2 D_m / n$?
Première idée :

$$\mathrm{crit}(m) = \frac{1}{n} \left\| Y - \widehat{F}_m \right\|^2 + \frac{2\widehat{\sigma}_{m_0}^2 D_m}{n}$$

Défauts : il faut connaître/choisir m_0, surpénalisation d'un facteur inconnu

FPE (Akaike, 1970) et GCV (Craven & Wahba, 1979)

$$\mathrm{crit}_{\mathrm{FPE}}(m) = \frac{1}{n}\left\|Y - \widehat{F}_m\right\|^2 + \frac{2\widehat{\sigma}_m^2 D_m}{n} = \frac{1}{n}\left\|Y - \widehat{F}_m\right\|^2 \left(1 + \frac{2D_m}{n - D_m}\right)$$

(Akaike, 1970 ; voir aussi Baraud, Giraud & Huet, 2009)

FPE (Akaike, 1970) et GCV (Craven & Wahba, 1979)

$$\mathrm{crit}_{\mathrm{FPE}}(m) = \frac{1}{n}\left\|Y - \widehat{F}_m\right\|^2 + \frac{2\widehat{\sigma}_m^2 D_m}{n} = \frac{1}{n}\left\|Y - \widehat{F}_m\right\|^2 \left(1 + \frac{2D_m}{n - D_m}\right)$$

(Akaike, 1970 ; voir aussi Baraud, Giraud & Huet, 2009)

Validation croisée généralisée (GCV, Craven & Wahba, 1979)

$$\mathrm{crit}_{\mathrm{GCV}}(m) = \frac{1}{n} \frac{\left\|Y - \widehat{F}_m\right\|^2}{\left(1 - \frac{D_m}{n}\right)^2}$$

FPE (Akaike, 1970) et GCV (Craven & Wahba, 1979)

$$\mathrm{crit}_{\mathrm{FPE}}(m) = \frac{1}{n}\left\|Y - \widehat{F}_m\right\|^2 + \frac{2\widehat{\sigma}_m^2 D_m}{n} = \frac{1}{n}\left\|Y - \widehat{F}_m\right\|^2 \left(1 + \frac{2D_m}{n - D_m}\right)$$

(Akaike, 1970 ; voir aussi Baraud, Giraud & Huet, 2009)

Validation croisée généralisée (GCV, Craven & Wahba, 1979)

$$\mathrm{crit}_{\mathrm{GCV}}(m) = \frac{1}{n} \frac{\left\|Y - \widehat{F}_m\right\|^2}{\left(1 - \frac{D_m}{n}\right)^2}$$

Si $D_m \ll n$,

$$\mathrm{crit}_{\mathrm{GCV}}(m) \approx \frac{1}{n}\left\|Y - \widehat{F}_m\right\|^2 \frac{n + D_m}{n - D_m} = \frac{1}{n}\left\|Y - \widehat{F}_m\right\|^2 \left(1 + \frac{2D_m}{n - D_m}\right)$$

FPE (Akaike, 1970) et GCV (Craven & Wahba, 1979)

$$\mathrm{crit}_{\mathrm{FPE}}(m) = \frac{1}{n}\left\|Y - \widehat{F}_m\right\|^2 + \frac{2\widehat{\sigma}_m^2 D_m}{n} = \frac{1}{n}\left\|Y - \widehat{F}_m\right\|^2 \left(1 + \frac{2D_m}{n - D_m}\right)$$

(Akaike, 1970 ; voir aussi Baraud, Giraud & Huet, 2009)

Validation croisée généralisée (GCV, Craven & Wahba, 1979)

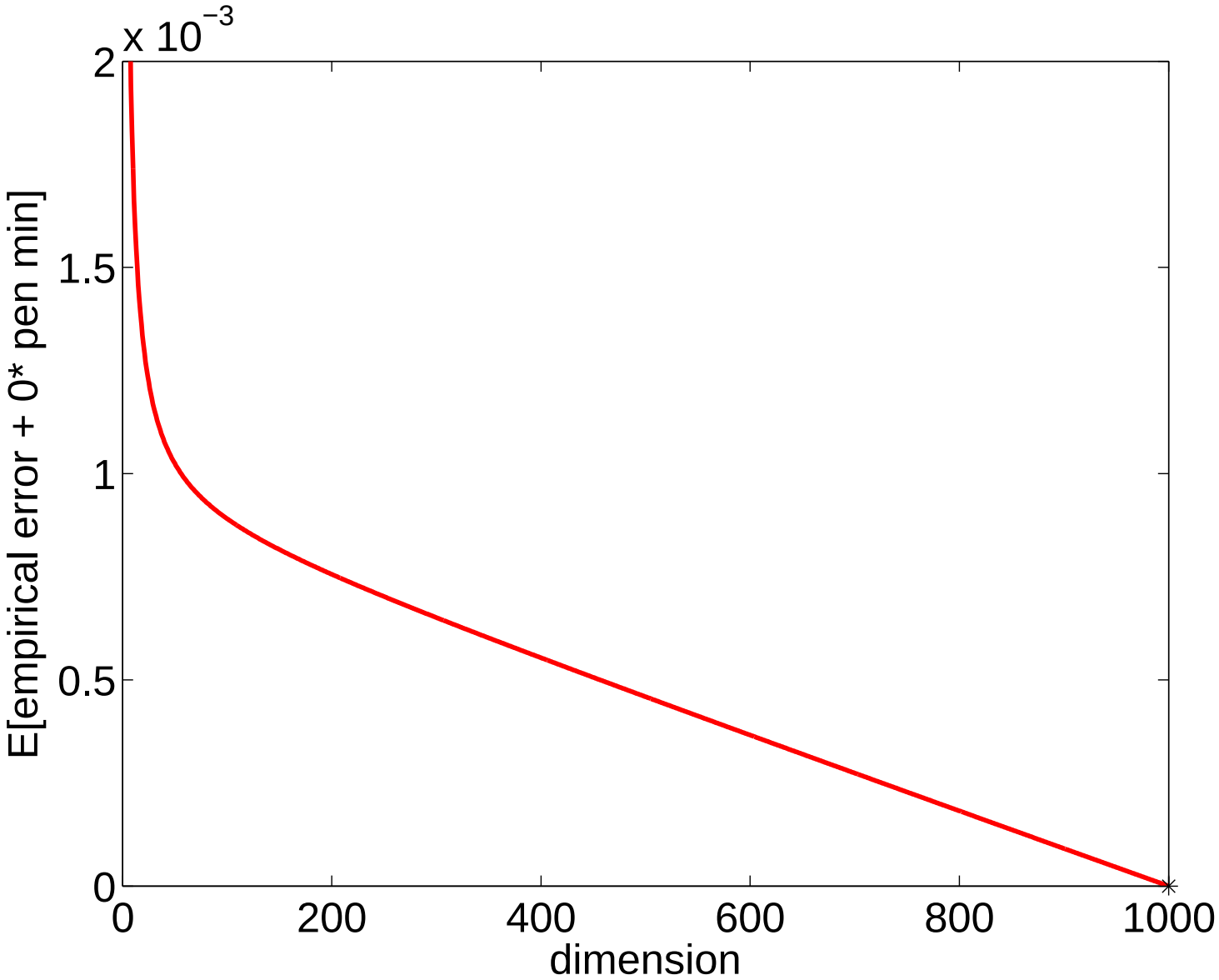
$$\mathrm{crit}_{\mathrm{GCV}}(m) = \frac{1}{n} \frac{\left\|Y - \widehat{F}_m\right\|^2}{\left(1 - \frac{D_m}{n}\right)^2}$$

Si $D_m \ll n$,

$$\mathrm{crit}_{\mathrm{GCV}}(m) \approx \frac{1}{n}\left\|Y - \widehat{F}_m\right\|^2 \frac{n + D_m}{n - D_m} = \frac{1}{n}\left\|Y - \widehat{F}_m\right\|^2 \left(1 + \frac{2D_m}{n - D_m}\right)$$

Défauts : pour les gros modèles $\left\|Y - \widehat{F}_m\right\|^2 \approx 0$

$$\mathbb{E}[\|Y - \widehat{F}_m\|^2] = n\sigma^2 + \|F - F_m\|^2 - \sigma^2 \times D_m$$



Plan

Pénalité minimale : heuristique

Pour tout $C > 0$,

$$\widehat{m}(C) \in \arg\min_{m \in \mathcal{M}_n} \left\{ \frac{1}{n} \left\| Y - \widehat{F}_m \right\|^2 + \frac{C D_m}{n} \right\}$$

$\Rightarrow \exists C_{\min}$ t.q. : pour $C < C_{\min}$, $\widehat{F}_{\widehat{m}(C)}$ sur-apprend

pour $C > C_{\min}$, inégalité-oracle pour $\widehat{F}_{\widehat{m}(C)}$?

Pénalité minimale : heuristique

Pour tout $C > 0$,

$$\widehat{m}(C) \in \arg\min_{m \in \mathcal{M}_n} \left\{ \frac{1}{n} \left\| Y - \widehat{F}_m \right\|^2 + \frac{C D_m}{n} \right\}$$

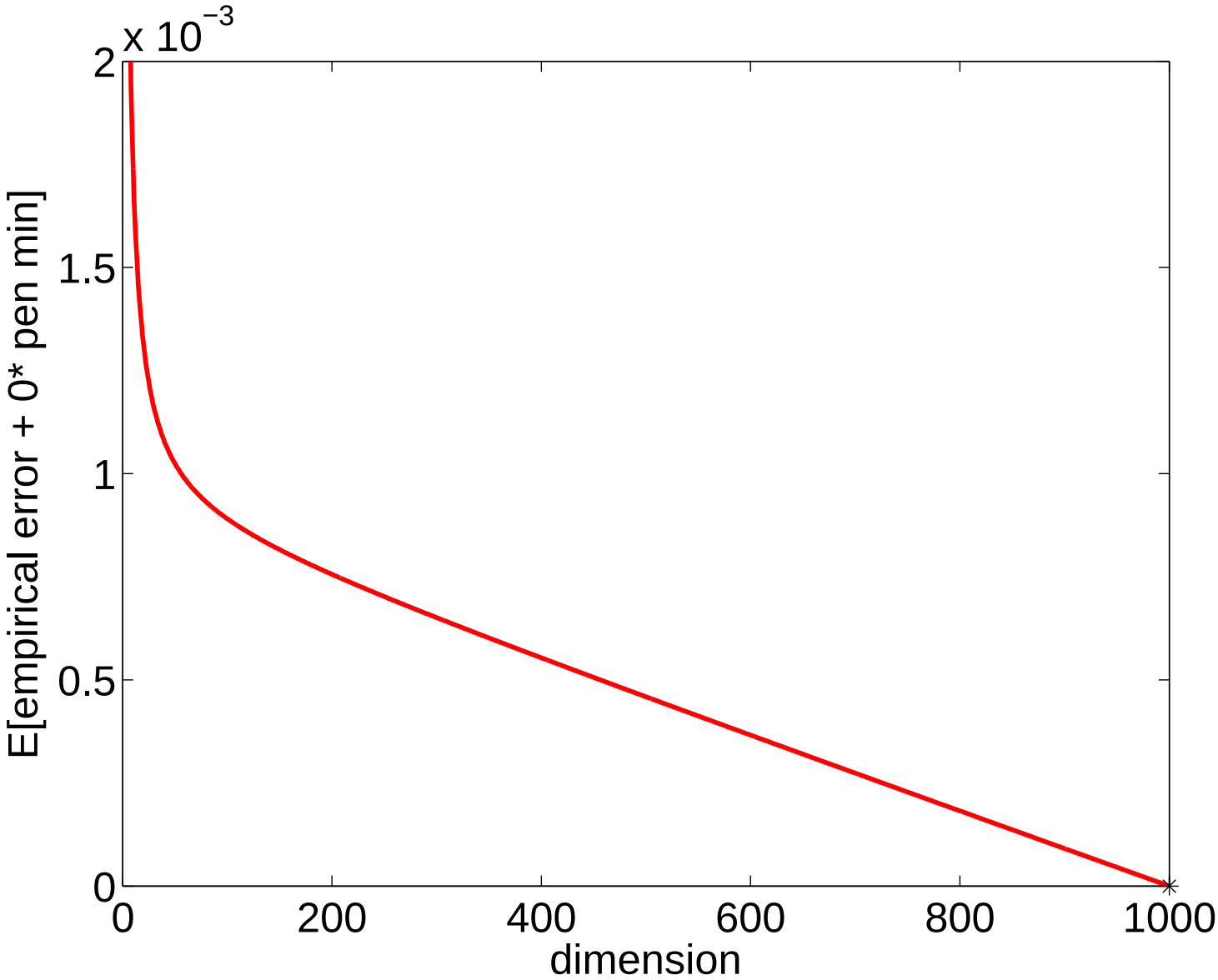
$\Rightarrow \exists C_{\min}$ t.q. : pour $C < C_{\min}$, $\widehat{F}_{\widehat{m}(C)}$ sur-apprend

pour $C > C_{\min}$, inégalité-oracle pour $\widehat{F}_{\widehat{m}(C)}$?

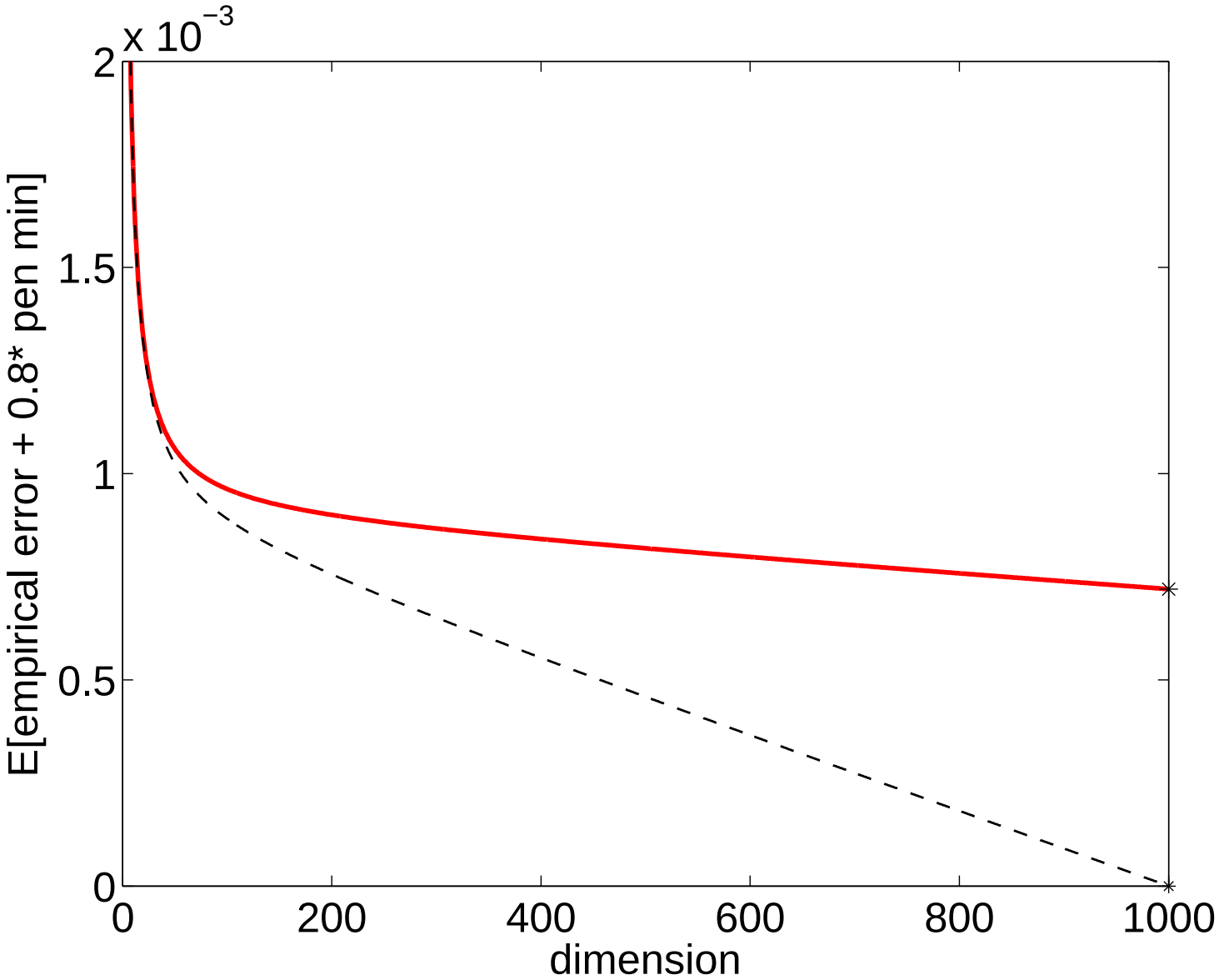
$$m^\star(C) \in \arg\min_{m \in \mathcal{M}_n} \left\{ \mathbb{E}\left[\frac{1}{n} \left\| Y - \widehat{F}_m \right\|^2 + \frac{C D_m}{n} \right] \right\}$$

$$= \arg\min_{m \in \mathcal{M}_n} \left\{ \frac{1}{n} \left\| F - F_m \right\|^2 + (C - \sigma^2) \frac{D_m}{n} \right\}$$

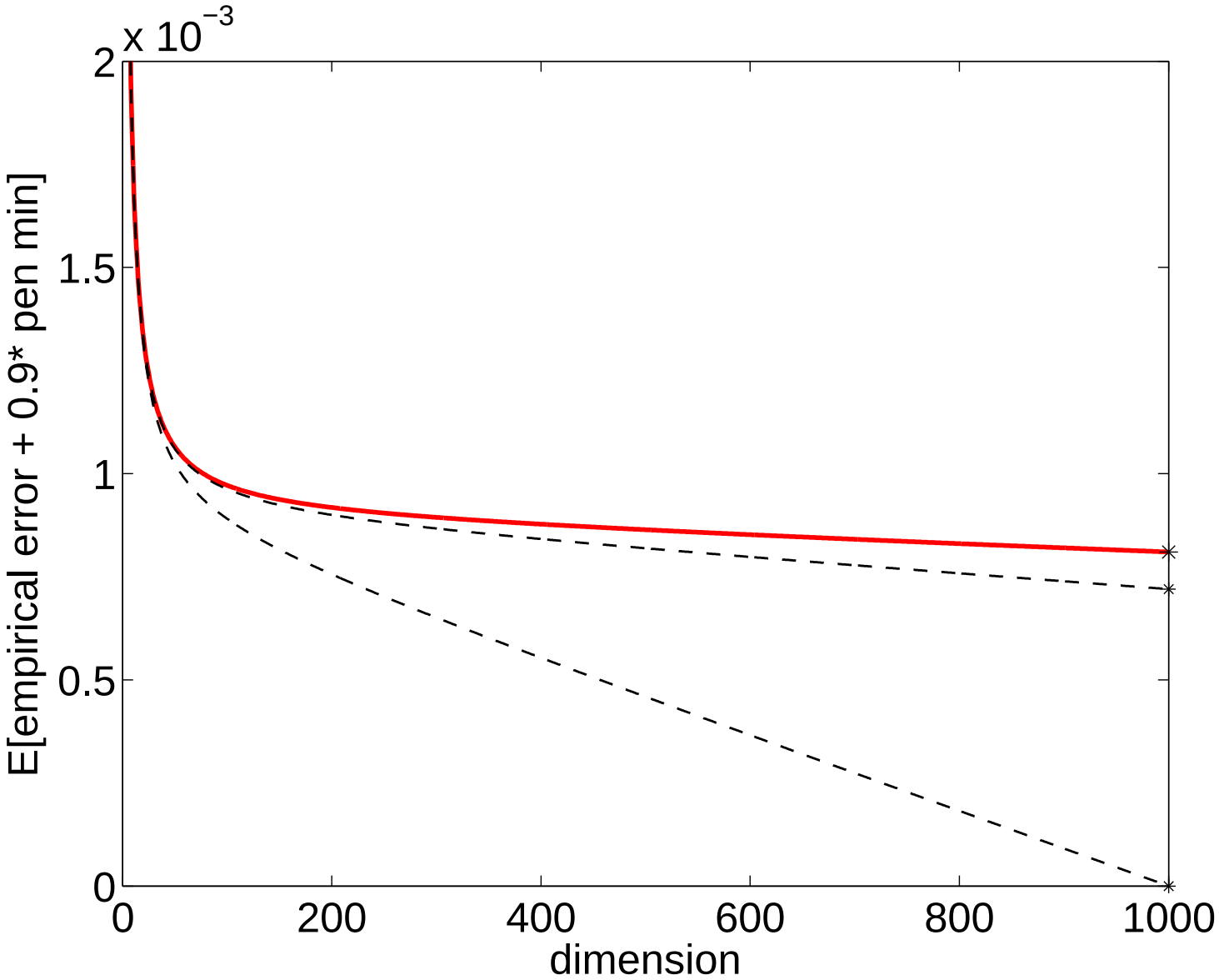
$\mathbb{E}[n^{-1}\|Y - \widehat{F}_m\|^2] + 0 \times \sigma^2 D_m n^{-1}$



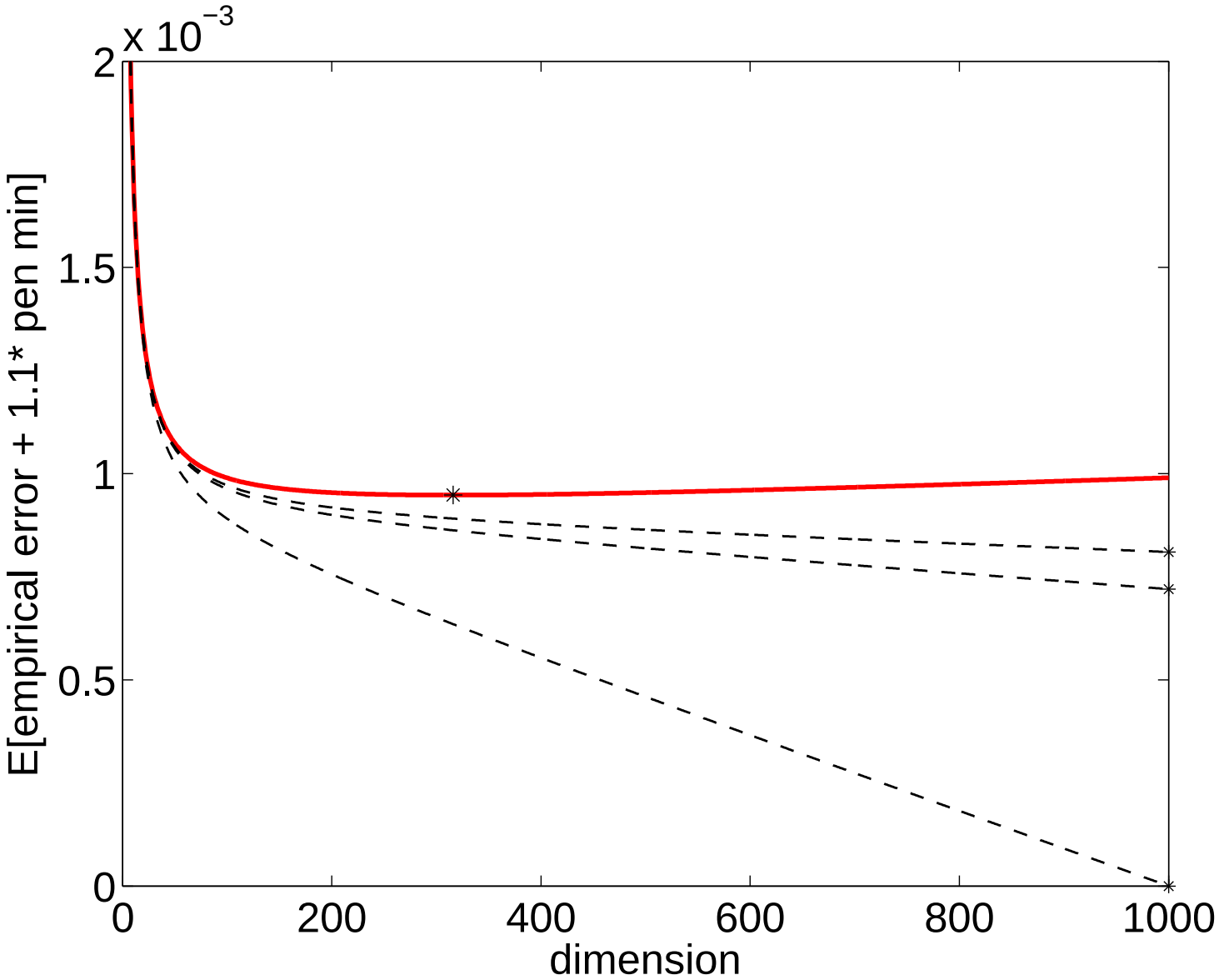
$\mathbb{E}[n^{-1}\|Y - \widehat{F}_m\|^2] + 0.8 \times \sigma^2 D_m n^{-1}$



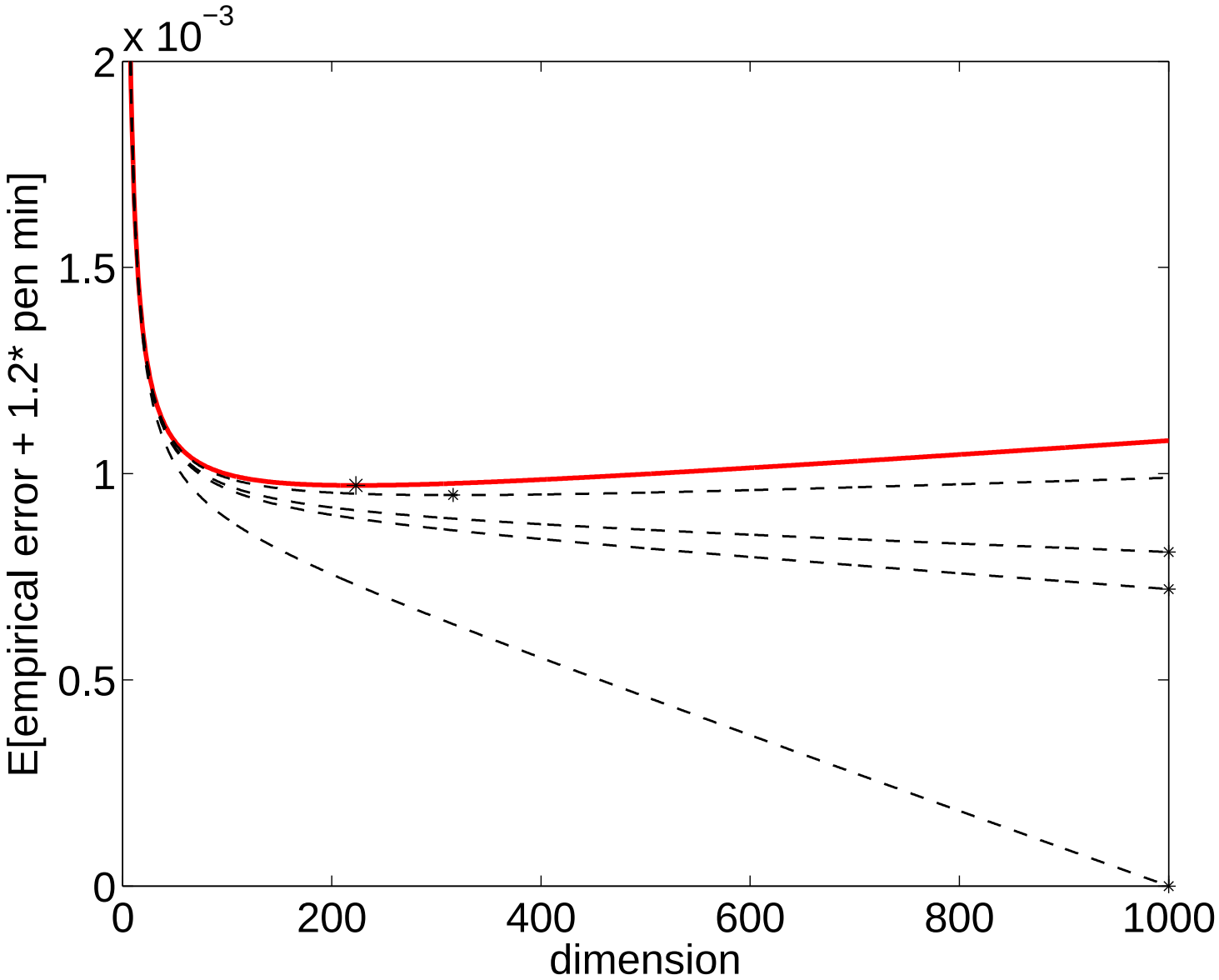
$\mathbb{E}[n^{-1}\|Y - \widehat{F}_m\|^2] + 0.9 \times \sigma^2 D_m n^{-1}$



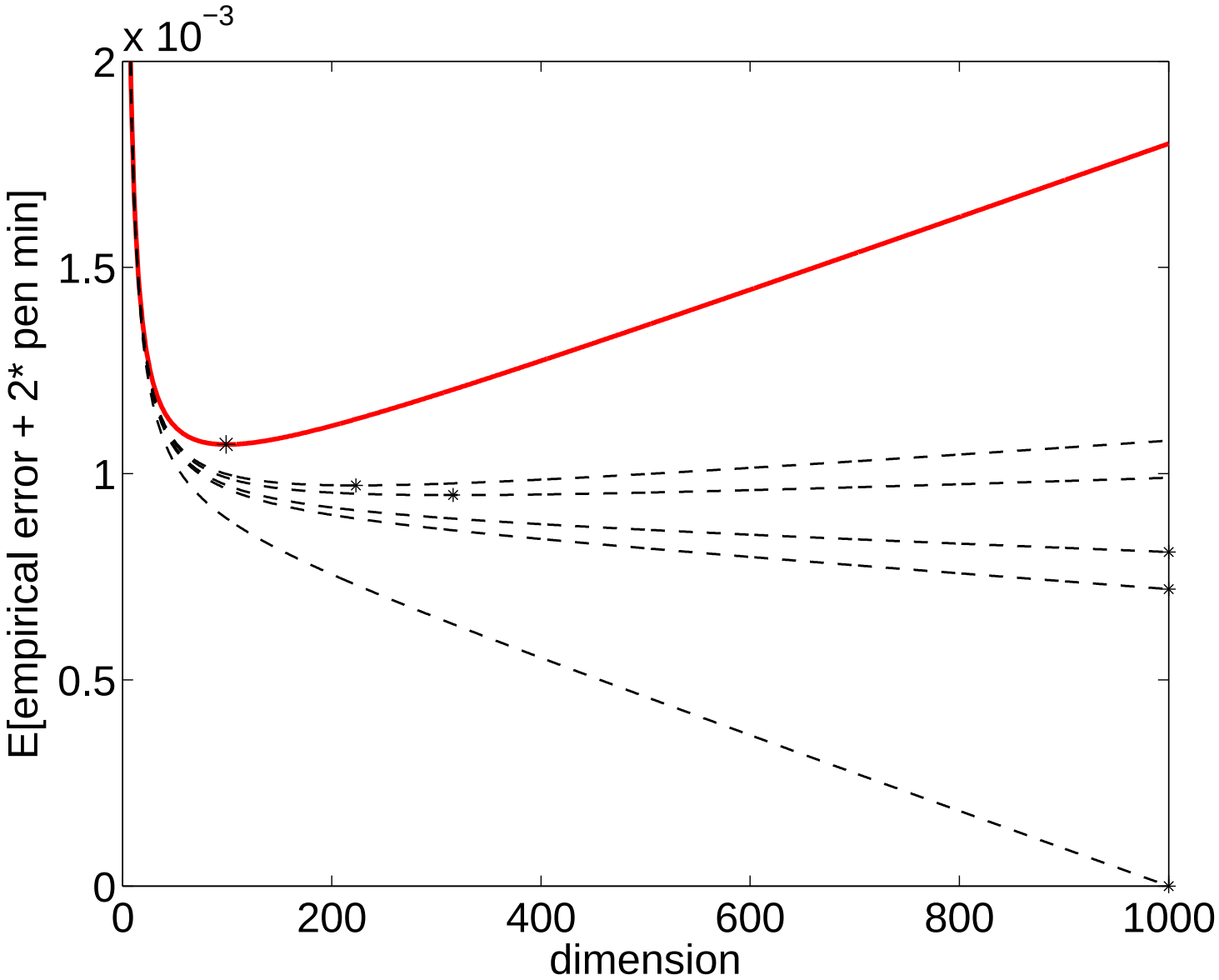
$\mathbb{E}[n^{-1}\|Y - \widehat{F}_m\|^2] + 1.1 \times \sigma^2 D_m n^{-1}$



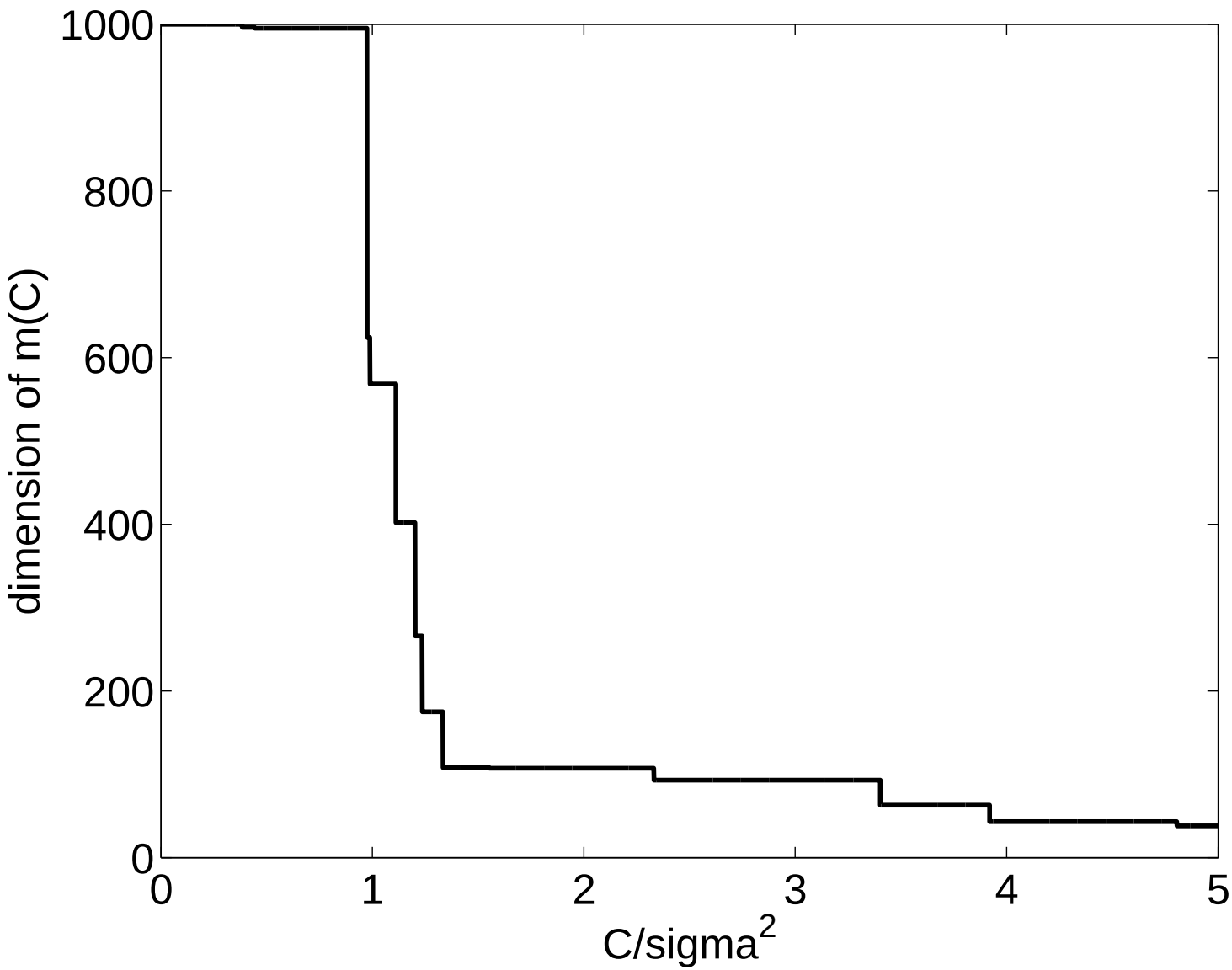
$\mathbb{E}[n^{-1}\|Y - \widehat{F}_m\|^2] + 1.2 \times \sigma^2 D_m n^{-1}$



$\mathbb{E}[n^{-1}\|Y - \widehat{F}_m\|^2] + 2 \times \sigma^2 D_m n^{-1}$



Saut de dimension



Calibration de pénalités (Birgé & Massart 2007)

1. pour tout $C > 0$, calculer

$$\widehat{m}(C) \in \arg\min_{m \in \mathcal{M}_n} \left\{ \frac{1}{n} \left\| Y - \widehat{F}_m \right\|^2 + C \frac{D_m}{n} \right\}$$

2. trouver $\widehat{C}_{\text{min}}$ tel que $D_{\widehat{m}(C)}$ est "très grande" lorsque $C < \widehat{C}_{\text{min}}$ et "raisonnablement petite" lorsque $C > \widehat{C}_{\text{min}}$

3. choisir $\widehat{m} = \widehat{m}\left(2\widehat{C}_{\text{min}}\right)$

Version rigoureuse : hypothèses et concentration

Hypothèses :

- complexité polynomiale : $\mathrm{Card}(\mathcal{M}_n) \leq C_{\mathcal{M}} n^{\alpha}$
- bruit Gaussien homoscédastique, design déterministe
- $\exists m_1,\, m_2 \in \mathcal{M}_n$ t.q. $D_{m_1} \geq n/2$, $D_{m_2} \leq \sqrt{n}$ et $\forall i \in \{1,2\}$, $n^{-1}\|F - F_{m_i}\|^2 \leq \sigma^2 \sqrt{\ln(n)/n}$

Proposition

Si $\xi \sim \mathcal{N}(0, I_n)$, $\alpha \in \mathbb{R}^n$, $M \in \mathcal{M}_n(\mathbb{R})$, *pour tout* $x \geq 0$,

$$\mathbb{P}\left(|\langle \xi,\, \alpha \rangle| \leq \sqrt{2x}\, \|\alpha\|_2 \right) \geq 1 - 2e^{-x}$$

$$\mathbb{P}\left(|\langle \xi,\, M\xi \rangle - \mathrm{tr}(M)| \leq 2\sqrt{x\,\mathrm{tr}\left(M^\top M\right) + 2\,|||M|||\,x} \right) \geq 1 - 2e^{-x}$$

Théorème (1) : Pénalité minimale / Saut de dimension

Théorème (Birgé & Massart 2007, A. & Bach 2009)

Avec probabilité au moins $1 - 4C_{\mathcal{M}} n^{-2}$, *si* $n \geq n_0(\alpha)$,

$$\forall C < \left(1 - 42\sqrt{\frac{(\alpha+2)\ln(n)}{n}}\right)\sigma^2, \quad D_{\widehat{m}(C)} \geq \frac{n}{3}$$

$$\forall C > \left(1 + 8\frac{\sqrt{(\alpha+2)\ln(n)}}{n^{1/4}}\right)\sigma^2, \quad D_{\widehat{m}(C)} \leq n^{3/4}$$

et dans le premier cas,

$$\left\| F - \widehat{F}_{\widehat{m}(C)} \right\|^2 \geq \ln(n) \inf_{m \in \mathcal{M}_n} \left\{ \left\| F - \widehat{F}_m \right\|^2 \right\}.$$

Théorème (2) : Inégalité-oracle

Théorème (Birgé & Massart 2007)

Pour tout $x \geq 0$, *avec probabilité* $1 - 4\,\mathrm{Card}(\mathcal{M}_n)e^{-x}$, *pour tout* $K > 1$, $\delta > 0$, *et tout*

$$\widehat{m}\left(K\sigma^2\right) \in \arg\min_{m \in \mathcal{M}_n} \left\{ \frac{1}{n} \left\| Y - \widehat{F}_m \right\|^2 + \frac{K\sigma^2 \dim(S_m)}{n} \right\} ,$$

$$\frac{1}{n} \left\| \widehat{F}_{\widehat{m}(K\sigma^2)} - F \right\|^2 \leq \left(\frac{1 + (K-2)_+}{1 - (2-K)_+} + \delta \right) \inf_{m \in \mathcal{M}_n} \left\{ \frac{1}{n} \left\| \widehat{F}_m - F \right\|^2 \right\} + \frac{C(K) x \sigma^2}{\delta n}$$

Théorème (2) : Inégalité-oracle

Théorème (Birgé & Massart 2007)

Si $\mathbb{P}(2\widehat{C} \in [(1-\eta_-)2\sigma^2; (1+\eta_+)2\sigma^2]) \geq 1 - 4C_{\mathcal{M}} n^{-2}$.

Pour tout $x \geq 0$, *avec probabilité* $1 - 4\,\mathrm{Card}(\mathcal{M}_n)e^{-x} - 4C_{\mathcal{M}} n^{-2}$, *pour tout* $\delta > 0$, *et tout*

$$\widehat{m}\left(2\widehat{C}\right) \in \arg\min_{m \in \mathcal{M}_n} \left\{ \frac{1}{n} \left\| Y - \widehat{F}_m \right\|^2 + \frac{2\widehat{C}\,\dim(S_m)}{n} \right\} ,$$

$$\frac{1}{n} \left\| \widehat{F}_{\widehat{m}(2\widehat{C})} - F \right\|^2 \leq \left(\frac{1+\eta_+}{1-\eta_-} + \delta \right) \inf_{m \in \mathcal{M}_n} \left\{ \frac{1}{n} \left\| \widehat{F}_m - F \right\|^2 \right\} + \frac{\max\{ C(2-\eta_-), C(2+\eta_+) \}\, x\sigma^2}{\delta n}$$

Théorème (2) : Inégalité-oracle

Théorème (Birgé & Massart 2007)

On prend $x = (\alpha + 2)\ln(n)$ *et on suppose* $n \geq n_0(\alpha)$.

Avec probabilité $1 - 4C_{\mathcal{M}} n^{-2}$, *pour tout*

$$\widehat{m}\left(2\widehat{C}\right) \in \arg\min_{m \in \mathcal{M}_n} \left\{ \frac{1}{n} \left\| Y - \widehat{F}_m \right\|^2 + \frac{2\widehat{C}\dim(S_m)}{n} \right\} ,$$

$$\frac{1}{n}\left\| \widehat{F}_{\widehat{m}(2\widehat{C})} - F \right\|^2 \leq \left(1 + \frac{L_\alpha \sqrt{\ln(n)}}{n^{1/4}} + \delta \right) \inf_{m \in \mathcal{M}_n} \left\{ \frac{1}{n}\left\| \widehat{F}_m - F \right\|^2 \right\} + \frac{L_\alpha \ln(n)\sigma^2}{\delta n}$$

Plan

Heuristique de pente (Birgé & Massart, 2007)

1. existence d'une pénalité minimale $\mathrm{pen}_{\min}(m)$:

$$\widehat{m}_{\min}(C) \in \arg\min_{m \in \mathcal{M}_n} \left\{ P_n \gamma\left(\widehat{s}_m\right) + C\,\mathrm{pen}_{\min}(m) \right\}$$

$$\frac{\ell\left(s^\star, \widehat{s}_{\widehat{m}_{\min}(C)}\right)}{\inf_{m \in \mathcal{M}_n} \left\{ \ell\left(s^\star, \widehat{s}_m\right) \right\}} \quad \text{saute en } C = 1$$

Heuristique de pente (Birgé & Massart, 2007)

1. existence d'une pénalité minimale $\mathrm{pen}_{\min}(m)$:

$$\widehat{m}_{\min}(C) \in \arg\min_{m \in \mathcal{M}_n} \left\{ P_n \gamma\left(\widehat{s}_m\right) + C\,\mathrm{pen}_{\min}(m) \right\}$$

$$\frac{\ell\left(s^\star, \widehat{s}_{\widehat{m}_{\min}(C)}\right)}{\inf_{m \in \mathcal{M}_n} \left\{ \ell\left(s^\star, \widehat{s}_m\right) \right\}} \quad \text{saute en } C = 1$$

2. pénalité minimale détectable :
   $\mathcal{C}_{\widehat{m}_{\min}(C)}$ "saute" au voisinage de $C = 1$

Heuristique de pente (Birgé & Massart, 2007)

1. existence d'une pénalité minimale $\mathrm{pen}_{\min}(m)$:

$$\widehat{m}_{\min}(C) \in \arg\min_{m \in \mathcal{M}_n} \left\{ P_n \gamma\left(\widehat{s}_m\right) + C\,\mathrm{pen}_{\min}(m) \right\}$$

$$\frac{\ell\left(s^\star, \widehat{s}_{\widehat{m}_{\min}(C)}\right)}{\inf_{m \in \mathcal{M}_n} \left\{ \ell\left(s^\star, \widehat{s}_m\right) \right\}} \quad \text{saute en } C = 1$$

2. pénalité minimale détectable :
   $\mathcal{C}_{\widehat{m}_{\min}(C)}$ "saute" au voisinage de $C = 1$

3. lien entre pénalités minimales et optimales :

$$\mathrm{pen}_{\mathrm{opt}}(m) \approx 2\,\mathrm{pen}_{\min}(m)$$

Calibration par heuristique de pente

$$\mathrm{pen}_0(m) \propto \mathrm{pen}_{\mathrm{min}}(m) \qquad (\mathcal{C}_m)_{m \in \mathcal{M}_n}$$

1. pour tout $C > 0$, calculer

$$\widehat{m}(C) \in \arg\min_{m \in \mathcal{M}_n} \left\{ P_n \gamma\left(\widehat{s}_m\right) + C\, \mathrm{pen}_0(m) \right\} \quad ,$$

2. trouver $\widehat{C}_{\mathrm{min}}$ tel que $\mathcal{C}_{\widehat{m}(C)}$ est "très grande" lorsque $C < \widehat{C}_{\mathrm{min}}$ et "raisonnablement petite" lorsque $C > \widehat{C}_{\mathrm{min}}$,

3. choisir $\widehat{m} = \widehat{m}\left(2\widehat{C}_{\mathrm{min}}\right)$.

Ingrédients de l'heuristique de pente

$$p_2(m) = P_n\left(\gamma\left(s_m^\star\right) - \gamma\left(\widehat{s}_m\right)\right) \qquad \mathrm{pen}_{\mathrm{min}}(m) = \mathbb{E}\left[p_2(m)\right]$$

Ingrédients de l'heuristique de pente

$$p_2(m) = P_n\left(\gamma\left(s_m^\star\right) - \gamma\left(\widehat{s}_m\right)\right) \qquad \mathrm{pen}_{\min}(m) = \mathbb{E}\left[p_2(m)\right]$$

$$p_1(m) = P\left(\gamma\left(\widehat{s}_m\right) - \gamma\left(s_m^\star\right)\right) \qquad \delta(m) = (P - P_n)\gamma\left(s_m^\star\right)$$

$$\mathrm{pen}_{\mathrm{id}}(m) = p_1(m) + p_2(m) - \delta(m)$$
$$\mathbb{E}\left[\mathrm{pen}_{\mathrm{id}}(m)\right] = \mathrm{pen}_{\mathrm{opt}}(m) = \mathbb{E}\left[p_1(m)\right] + \mathbb{E}\left[p_2(m)\right]$$

Ingrédients de l'heuristique de pente

$$p_2(m) = P_n\left(\gamma\left(s_m^\star\right) - \gamma\left(\widehat{s}_m\right)\right) \qquad \mathrm{pen}_{\min}(m) = \mathbb{E}\left[p_2(m)\right]$$

$$p_1(m) = P\left(\gamma\left(\widehat{s}_m\right) - \gamma\left(s_m^\star\right)\right) \qquad \delta(m) = (P - P_n)\gamma\left(s_m^\star\right)$$

$$\mathrm{pen}_{\mathrm{id}}(m) = p_1(m) + p_2(m) - \delta(m)$$
$$\mathbb{E}\left[\mathrm{pen}_{\mathrm{id}}(m)\right] = \mathrm{pen}_{\mathrm{opt}}(m) = \mathbb{E}\left[p_1(m)\right] + \mathbb{E}\left[p_2(m)\right]$$

Heuristique : $p_1(m) \approx p_2(m)$

- concentrations p_1, p_2, δ
- $\mathbb{E}\left[p_1(m)\right] \approx \mathbb{E}\left[p_2(m)\right]$
- croissance de l'espérance pour compenser le biais

Résultats mathématiques existants

- Moindres carrés, régression homoscédastique Gaussienne (Birgé & Massart, 2007)
- Régressogrammes hétéroscedastiques (A. & Massart, 2009)
- Estimation de densité par moindres carrés, données i.i.d. (Lerasle, 2009) ou mélangeantes (Lerasle, 2010)
- Estimateurs par minimum de contraste régulier (Saumard, 2010)

Minimum de contraste régulier (Saumard, 2010)

- Contraste régulier sur un modèle S_m convexe :
  - $s_m^\star \in \arg\min_{t \in S_m} P\gamma(t)$ existe
  - $t \in S_m \mapsto P\gamma(t)$ strictement convexe
  - $\exists c > 0,\ t \in B_\infty(s_m^\star, c) \mapsto \gamma(t;\cdot) \in L_\infty(P)$ est $\mathcal{C}^3$

- Concentration de $p_1(m)$ et $p_2(m)$ autour de la même quantité déterministe $D_m \mathcal{K}_m^2/(4n)$ (non-observable a priori)

$+$ contrôle de $\|\widehat{s}_m - s_m^\star\|_\infty$

$\Rightarrow$ validation de l'heuristique de pente pour :
  - régression hétéroscédastique (histogrammes, polynômes par morceaux)
  - estimation de densité par moindres carrés
  - estimation de densité par maximum de vraisemblance sur des histogrammes

Résultats empiriques

- Détection de ruptures (Lebarbier, 2005)
- Modèles de mélanges Gaussiens (Maugis & Michel, 2008)
- Classification non-supervisée (choix du nombre de classes) (Baudry, 2009)
- Géométrie computationnelle (Caillerie & Michel, 2009)
- Lasso (Connault, 2011)
- ...

pour une liste plus complète, cf. Baudry, Maugis & Michel, 2010

Plan

Estimateurs linéaires en régression

- Définition :

$$\widehat{F} = AY$$

avec A déterministe (donc, A peut dépendre de X)

Estimateurs linéaires en régression

- Définition :

$$\widehat{F} = AY$$

avec A déterministe (donc, A peut dépendre de X)

- Exemple : Estimateur des moindres carrés : A matrice de projection orthogonale
- Exemple : Estimateur "régularisé" : A diagonalisable avec des valeurs propres $1 \geq \lambda_1 \geq \cdots \geq \lambda_n \geq 0$

Estimateurs linéaires en régression

- Définition :

$$\widehat{F} = AY$$

avec A déterministe (donc, A peut dépendre de X)

- Exemple : Estimateur des moindres carrés : A matrice de projection orthogonale
- Exemple : Estimateur "régularisé" : A diagonalisable avec des valeurs propres $1 \geq \lambda_1 \geq \cdots \geq \lambda_n \geq 0$
- Hypothèses :

$$\|A\| \leq 1 \qquad \text{ou} \qquad \|A\| \leq B$$

$$\mathrm{tr}(A^\top A) \leq \mathrm{tr}(A)$$

Régression ridge (1/2)

- Idée : chercher un estimateur ayant un risque empirique petit et une petite norme dans un espace fonctionnel $\mathcal{F}$

- $\mathcal{F} \subset \mathbb{S}$ est l'espace de Hilbert à noyau reproduisant (RKHS) associé à un noyau défini positif $k : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$

$$\widehat{f} \in \arg\min_{f \in \mathcal{F}} \left\{ \frac{1}{n} \sum_{i=1}^{n} (Y_i - f(X_i))^2 + \mu \| f \|_{\mathcal{F}}^2 \right\}$$

- Théorème du représentant $\Rightarrow \exists \widehat{\alpha} \in \mathbb{R}^n$ tel que $\widehat{f} = \sum_{i=1}^{n} \widehat{\alpha}_i k(X_i, \cdot)$

Régression ridge (2/2)

$$\widehat{f} = \sum_{i=1}^{n} \widehat{\alpha}_i k(X_i, \cdot) \in \arg\min_{f \in \mathcal{F}} \left\{ \frac{1}{n} \sum_{i=1}^{n} (Y_i - f(X_i))^2 + \mu \| f \|_{\mathcal{F}}^2 \right\}$$

- Design fixe : $\widehat{F} = (\widehat{f}(x_i))_{1 \leq i \leq n} = K\widehat{\alpha}$ avec $K = (k(x_i, x_j))_{1 \leq i,j \leq n}$ et

$$\widehat{\alpha} \in \arg\min_{\alpha \in \mathbb{R}^n} \left\{ \frac{1}{n} \| Y - K\alpha \|^2 + \mu \alpha^\top K \alpha \right\}$$

$$\Rightarrow \quad \widehat{F} = K(K + n\mu I_n)^{-1} Y$$

Régression ridge (2/2)

- Design fixe : $\widehat{F} = (\widehat{f}(x_i))_{1 \leq i \leq n} = K\widehat{\alpha}$ avec $K = (k(x_i, x_j))_{1 \leq i,j \leq n}$ et

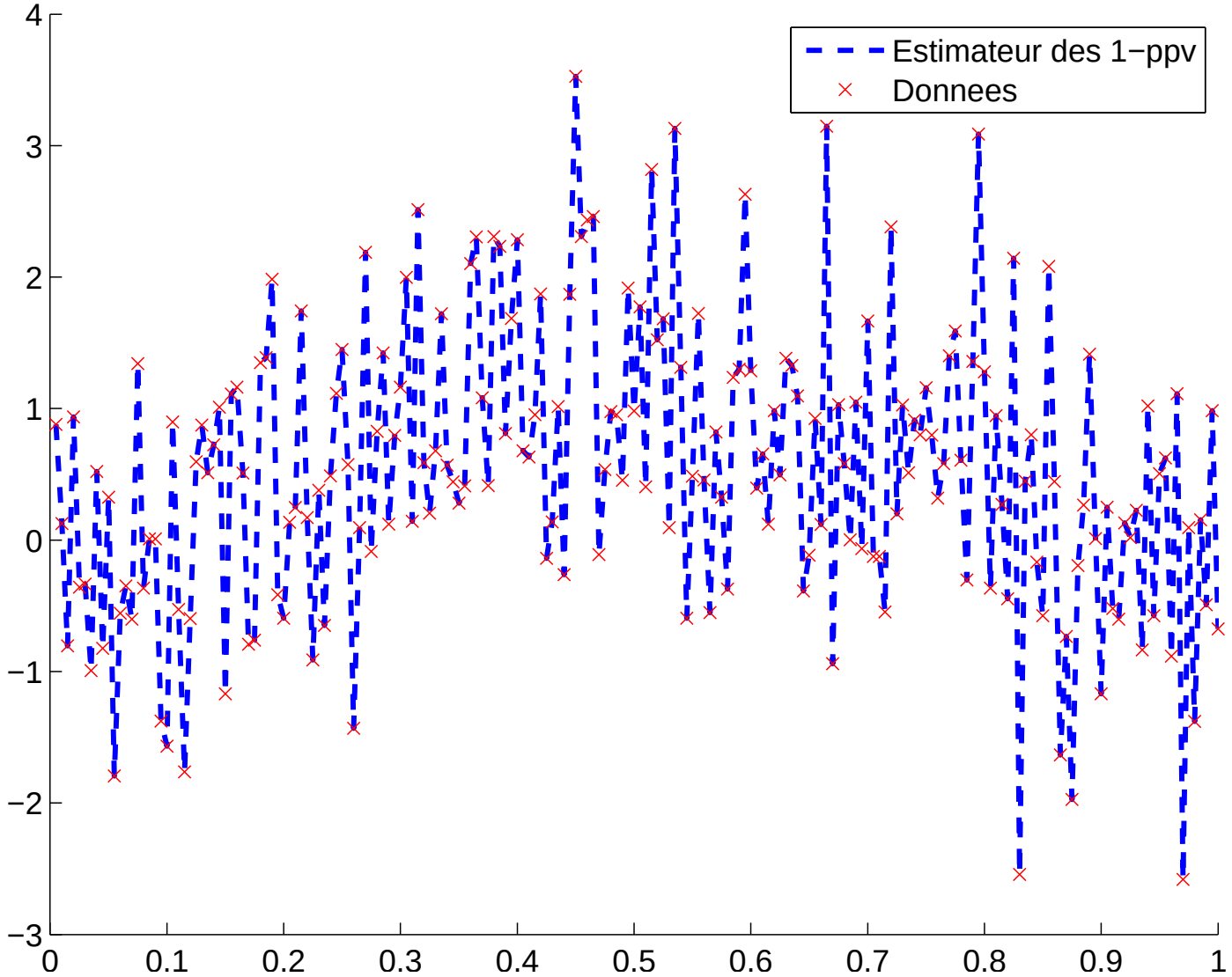
$$\widehat{\alpha} \in \arg\min_{\alpha \in \mathbb{R}^n} \left\{ \frac{1}{n} \|Y - K\alpha\|^2 + \mu \alpha^\top K \alpha \right\}$$

$$\Rightarrow \quad \widehat{F} = K(K + n\mu I_n)^{-1} Y$$

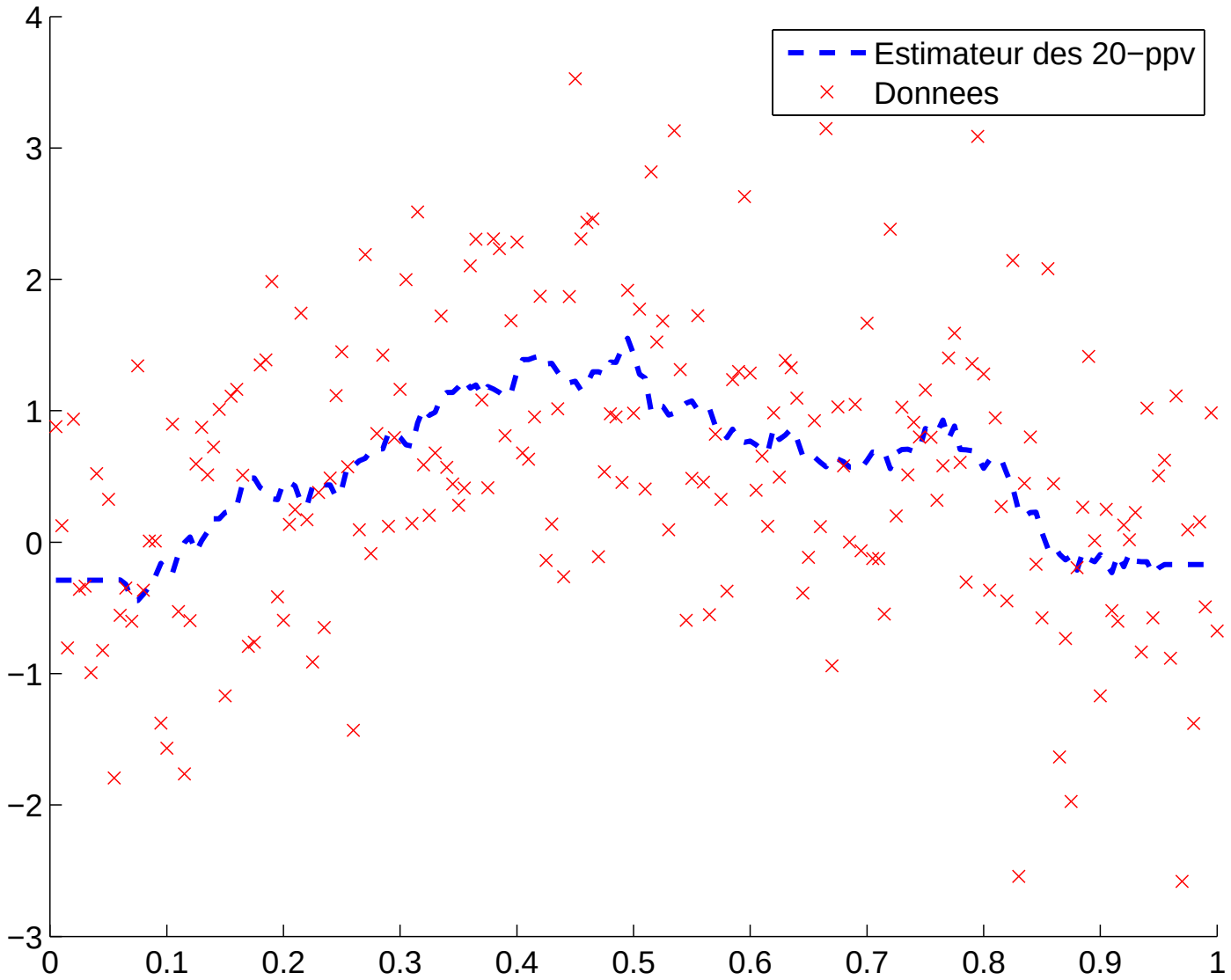
- Si $(e_j)_{1 \leq j \leq n}$ b.o.n. de vecteurs propres de K (valeurs propres $(\kappa_j)_{1 \leq j \leq n}$) :

$$\widehat{F} e_j = \frac{\kappa_j}{\kappa_j + n\mu I_n} e_j$$

Estimateur du plus proche voisin : $\widehat{F} = Y$



Estimateur des k-plus proches voisins ($k = 20$)



Estimateur des k-plus proches voisins

Pour tout $i \in \{1, \ldots, n\}$,

$$\widehat{F}_i = \frac{1}{k} \sum_{x_j \in \{ k\text{-ppv de } x_i \}} Y_j$$

$$= \sum_{1 \leq j \leq n} \left(\frac{1}{k} \mathbb{1}_{x_j \in \{ k\text{-ppv de } x_i \}} Y_j \right)$$

Estimateur des k-plus proches voisins

Pour tout $i \in \{1, \ldots, n\}$,

$$\widehat{F}_i = \frac{1}{k} \sum_{x_j \in \{ k\text{-ppv de } x_i \}} Y_j$$

$$= \sum_{1 \leq j \leq n} \left(\frac{1}{k} \mathbb{1}_{x_j \in \{ k\text{-ppv de } x_i \}} Y_j \right)$$

$$\Rightarrow \qquad \widehat{F} = A(k)\, Y$$

$$\text{avec} \quad A(k)_{i,j} = \frac{1}{k} \mathbb{1}_{x_j \in \{ k\text{-ppv de } x_i \}}$$

Estimateur de Nadaraya-Watson

Soit un "noyau" $K : \mathcal{X} \times \mathcal{X} \mapsto [0, +\infty)$.

$$\forall i, j \in \{1, \ldots, n\} , \quad A(K)_{i,j} = \frac{K(x_i, x_j)}{\sum_{1 \leq \ell \leq n} K(x_i, x_\ell)}$$

$$\widehat{F} = A(K) Y$$

Estimateur de Nadaraya-Watson

Soit un "noyau" $K : \mathcal{X} \times \mathcal{X} \mapsto [0, +\infty)$.

$$\forall i, j \in \{1, \ldots, n\}\ , \quad A(K)_{i,j} = \frac{K(x_i, x_j)}{\sum_{1 \leq \ell \leq n} K(x_i, x_\ell)}$$

$$\widehat{F} = A(K)\, Y$$

- Typiquement, $K(x, y) = g(d(x, y)/h)$ pour une distance d sur $\mathcal{X}$, une largeur de bande $h > 0$ et une fonction $g : [0, +\infty) \mapsto [0, +\infty)$ décroissante.

Estimateur de Nadaraya-Watson

Soit un "noyau" $K : \mathcal{X} \times \mathcal{X} \mapsto [0, +\infty)$.

$$\forall i, j \in \{1, \ldots, n\}, \quad A(K)_{i,j} = \frac{K(x_i, x_j)}{\sum_{1 \leq \ell \leq n} K(x_i, x_\ell)}$$

$$\widehat{F} = A(K)Y$$

- Typiquement, $K(x, y) = g(d(x, y)/h)$ pour une distance d sur $\mathcal{X}$, une largeur de bande $h > 0$ et une fonction $g : [0, +\infty) \mapsto [0, +\infty)$ décroissante.
- Noyau Gaussien $g(t) = \exp(-t^2)$.
- Noyau Fenêtre $g(t) = \mathbb{1}_{t \in [0,1]}$

Risque, risque empirique, pénalité idéale

$$Y = F + \varepsilon \quad \text{avec} \quad \mathbb{E}\left[\varepsilon_i^2\right] = \sigma^2$$

$$\widehat{F}_m = A_m Y$$

Risque, risque empirique, pénalité idéale

$$Y = F + \varepsilon \quad \text{avec} \quad \mathbb{E}\left[\varepsilon_i^2\right] = \sigma^2$$

$$\widehat{F}_m = A_m Y$$

$$\mathbb{E}\left[\frac{1}{n}\left\|\widehat{F}_m - F\right\|^2\right] = \frac{1}{n}\left\|(A_m - I)F\right\|^2 + \frac{\sigma^2 \operatorname{tr}(A_m^\top A_m)}{n}$$

Risque, risque empirique, pénalité idéale

$$Y = F + \varepsilon \quad \text{avec} \quad \mathbb{E}\left[\varepsilon_i^2\right] = \sigma^2$$

$$\widehat{F}_m = A_m Y$$

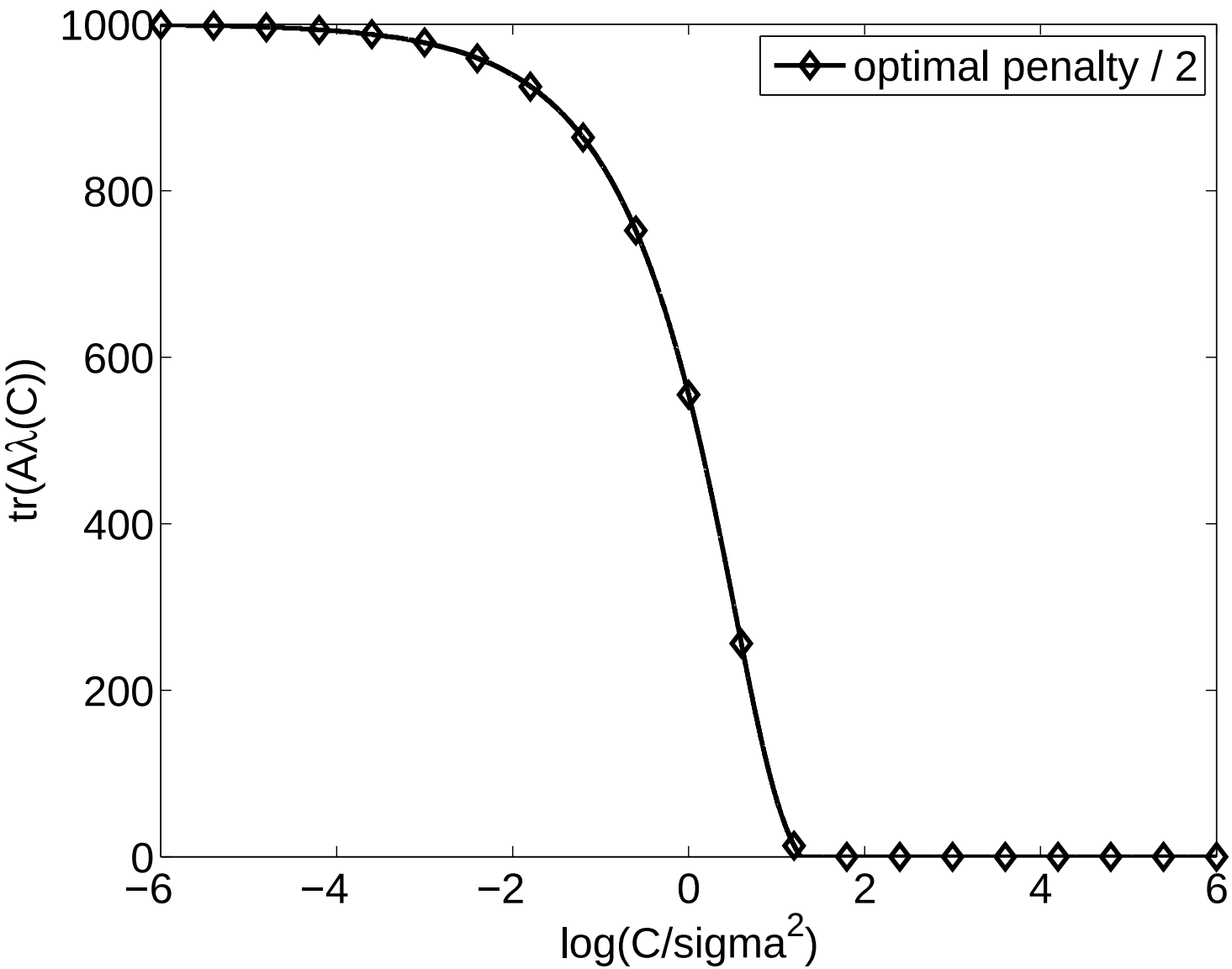
$$\mathbb{E}\left[\frac{1}{n}\left\|\widehat{F}_m - F\right\|^2\right] = \frac{1}{n}\left\|(A_m - I)F\right\|^2 + \frac{\sigma^2 \operatorname{tr}(A_m^\top A_m)}{n}$$

$$\operatorname{pen}_{\mathrm{id}}(m) = \frac{2}{n}\langle A_m\varepsilon,\, \varepsilon\rangle + \frac{2}{n}\langle (A_m - I_n)F,\, \varepsilon\rangle$$

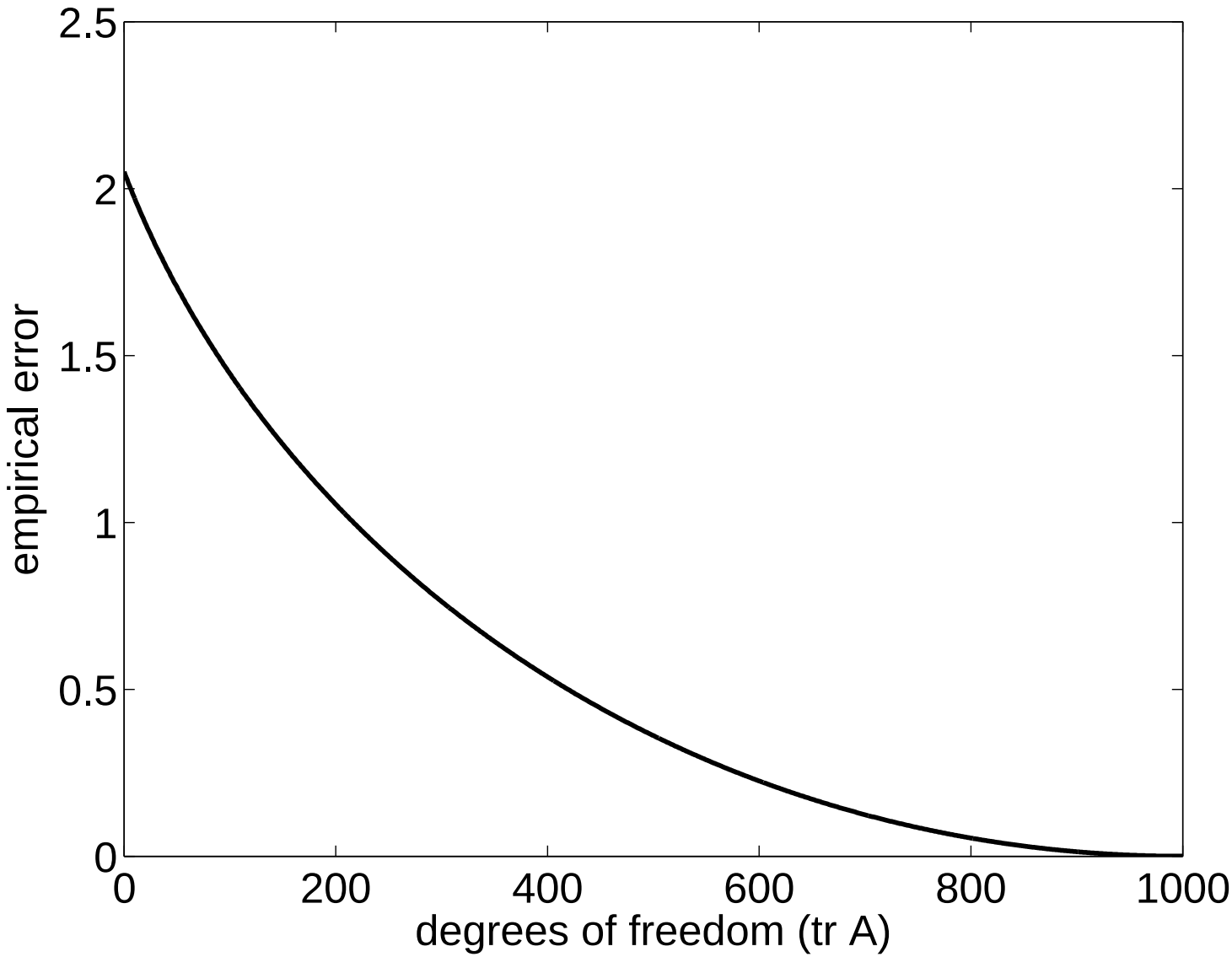
$$\mathbb{E}\left[\operatorname{pen}_{\mathrm{id}}(m)\right] = \frac{2\sigma^2 \operatorname{tr}(A_m)}{n} \quad \Rightarrow \quad C_L \text{ (Mallows, 1973)}$$

degrés de liberté généralisés : $\operatorname{tr}(A_m)$

Pas de saut de dimension avec une pénalité $\propto \mathrm{tr}(A_m)$



La pénalité minimale n'est pas proportionnelle à $\mathrm{tr}(A_m)$



Quelle pénalité minimale ?

$$
\begin{aligned}
p_2(m) &= \frac{1}{n}\left\|Y - F_m\right\|^2 - \frac{1}{n}\left\|Y - \widehat{F}_m\right\|^2 \\
&= \frac{2}{n}\left\langle \varepsilon,\, A_m\varepsilon \right\rangle - \frac{1}{n}\left\|A_m\varepsilon\right\|^2 - \frac{2}{n}\left\langle \varepsilon,\, A_m^\top (I_n - A_m) F \right\rangle
\end{aligned}
$$

Quelle pénalité minimale ?

$$
\begin{aligned}
p_2(m) &= \frac{1}{n}\left\|Y - F_m\right\|^2 - \frac{1}{n}\left\|Y - \widehat{F}_m\right\|^2 \\
&= \frac{2}{n}\left\langle \varepsilon,\, A_m\varepsilon \right\rangle - \frac{1}{n}\left\|A_m\varepsilon\right\|^2 - \frac{2}{n}\left\langle \varepsilon,\, A_m^\top (I_n - A_m) F \right\rangle
\end{aligned}
$$

donc

$$
\mathrm{pen}_{\mathrm{min}}(m) = \mathbb{E}\left[p_2(m)\right] = \frac{\sigma^2\left(2\,\mathrm{tr}(A_m) - \mathrm{tr}(A_m^\top A_m)\right)}{n}
$$

Quelle pénalité minimale ?

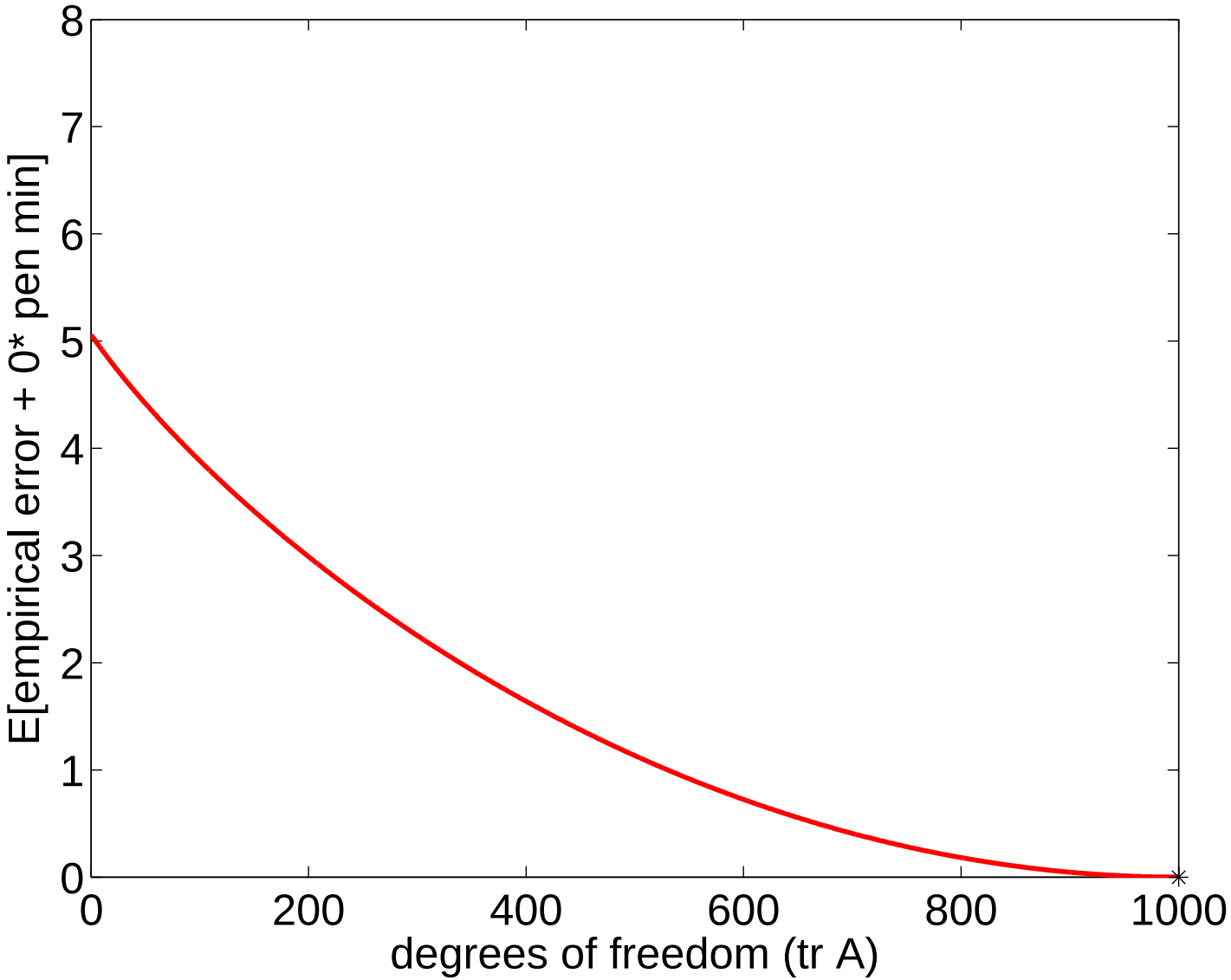
$$
\begin{aligned}
p_2(m) &= \frac{1}{n}\left\|Y - F_m\right\|^2 - \frac{1}{n}\left\|Y - \widehat{F}_m\right\|^2 \\
&= \frac{2}{n}\left\langle \varepsilon,\, A_m\varepsilon \right\rangle - \frac{1}{n}\left\|A_m\varepsilon\right\|^2 - \frac{2}{n}\left\langle \varepsilon,\, A_m^\top (I_n - A_m) F \right\rangle
\end{aligned}
$$

donc

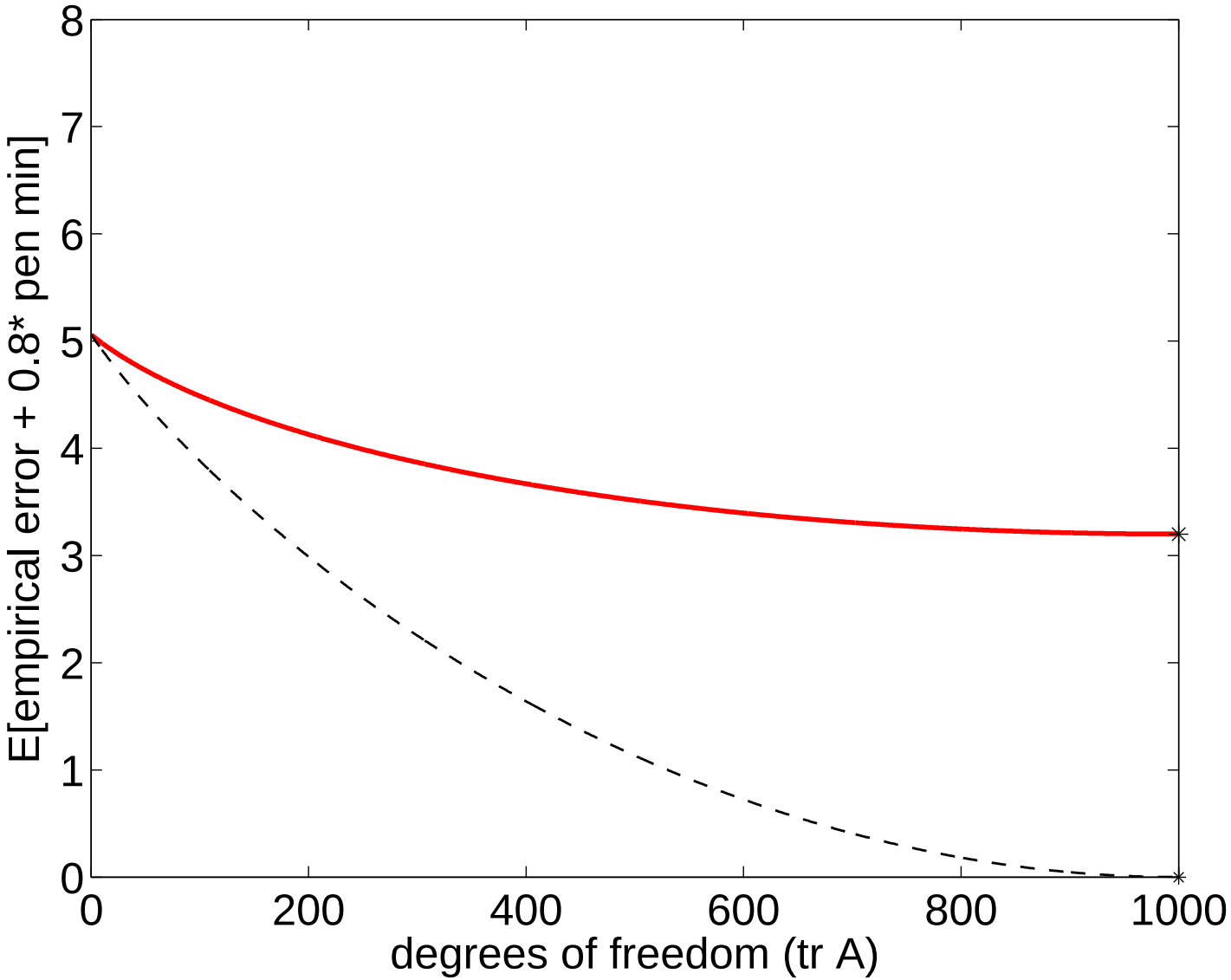
$$
\mathrm{pen}_{\mathrm{min}}(m) = \mathbb{E}\left[p_2(m)\right] = \frac{\sigma^2\left(2\,\mathrm{tr}(A_m) - \mathrm{tr}(A_m^\top A_m)\right)}{n}
$$

$$
\widehat{m}_{\mathrm{min}}(C) \in \arg\min_{m \in \mathcal{M}_n} \left\{ \frac{1}{n}\left\|Y - \widehat{F}_m\right\|^2 + \frac{C\left(2\,\mathrm{tr}(A_m) - \mathrm{tr}(A_m^\top A_m)\right)}{n} \right\}
$$

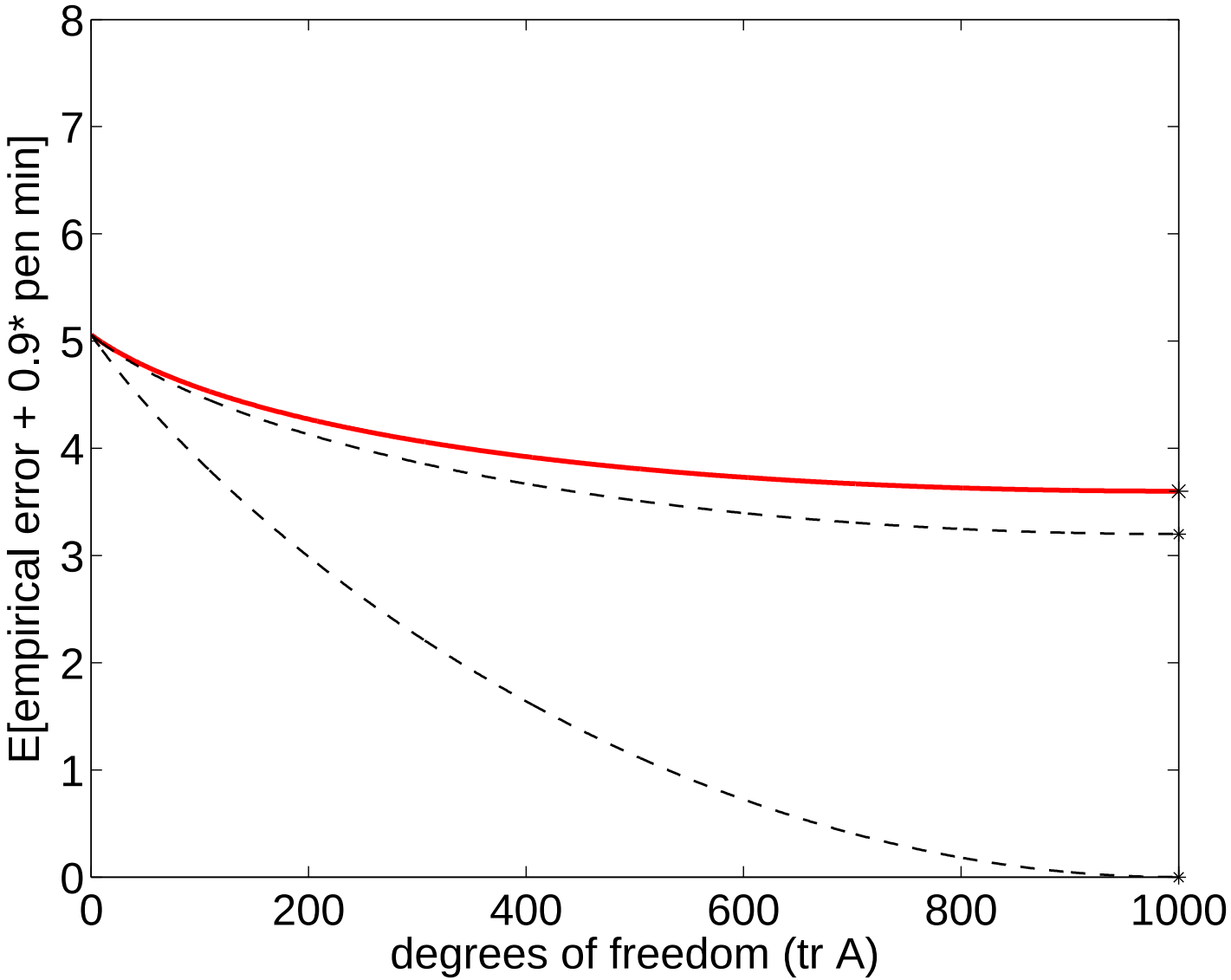
$$\mathbb{E}[n^{-1}\|Y - \widehat{F}_m\|^2] + 0 \times \sigma^2(2\,\mathrm{tr}(A_m) - \mathrm{tr}(A_m^\top A_m))n^{-1}$$



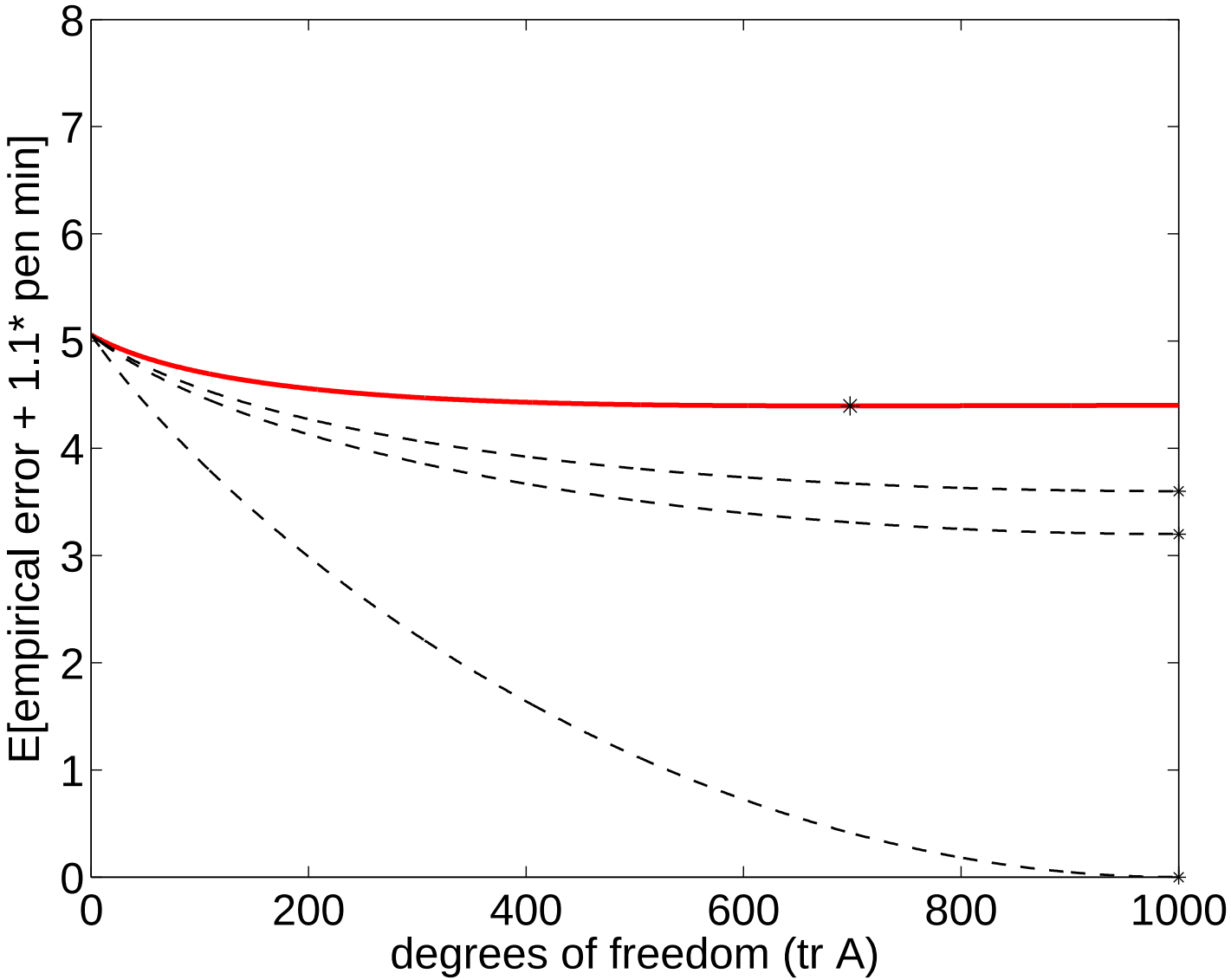
$$\mathbb{E}[n^{-1}\|Y - \widehat{F}_m\|^2] + 0.8 \times \sigma^2(2\,\mathrm{tr}(A_m) - \mathrm{tr}(A_m^\top A_m))n^{-1}$$



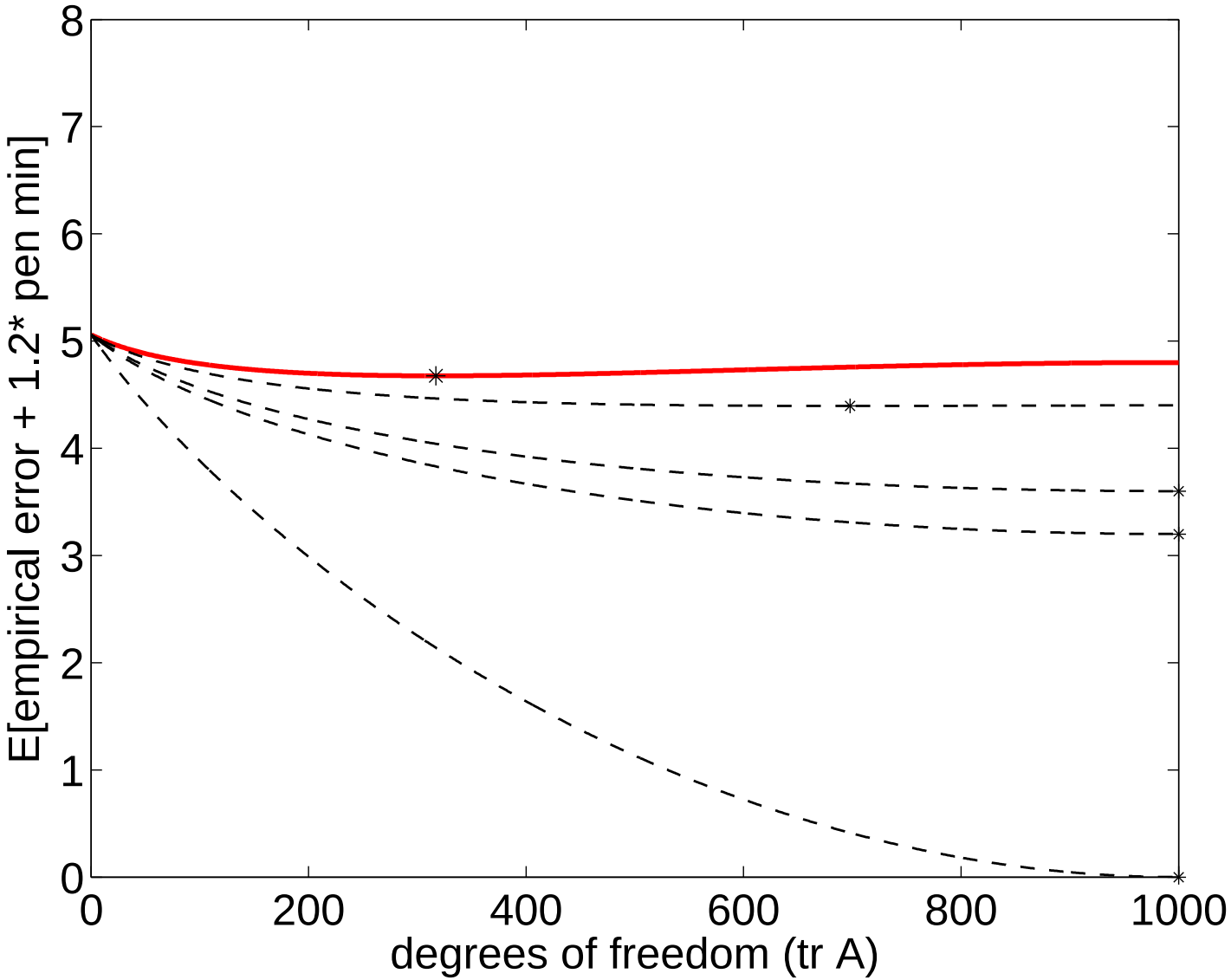
$$\mathbb{E}[n^{-1}\|Y - \widehat{F}_m\|^2] + 0.9 \times \sigma^2(2\,\mathrm{tr}(A_m) - \mathrm{tr}(A_m^\top A_m))n^{-1}$$



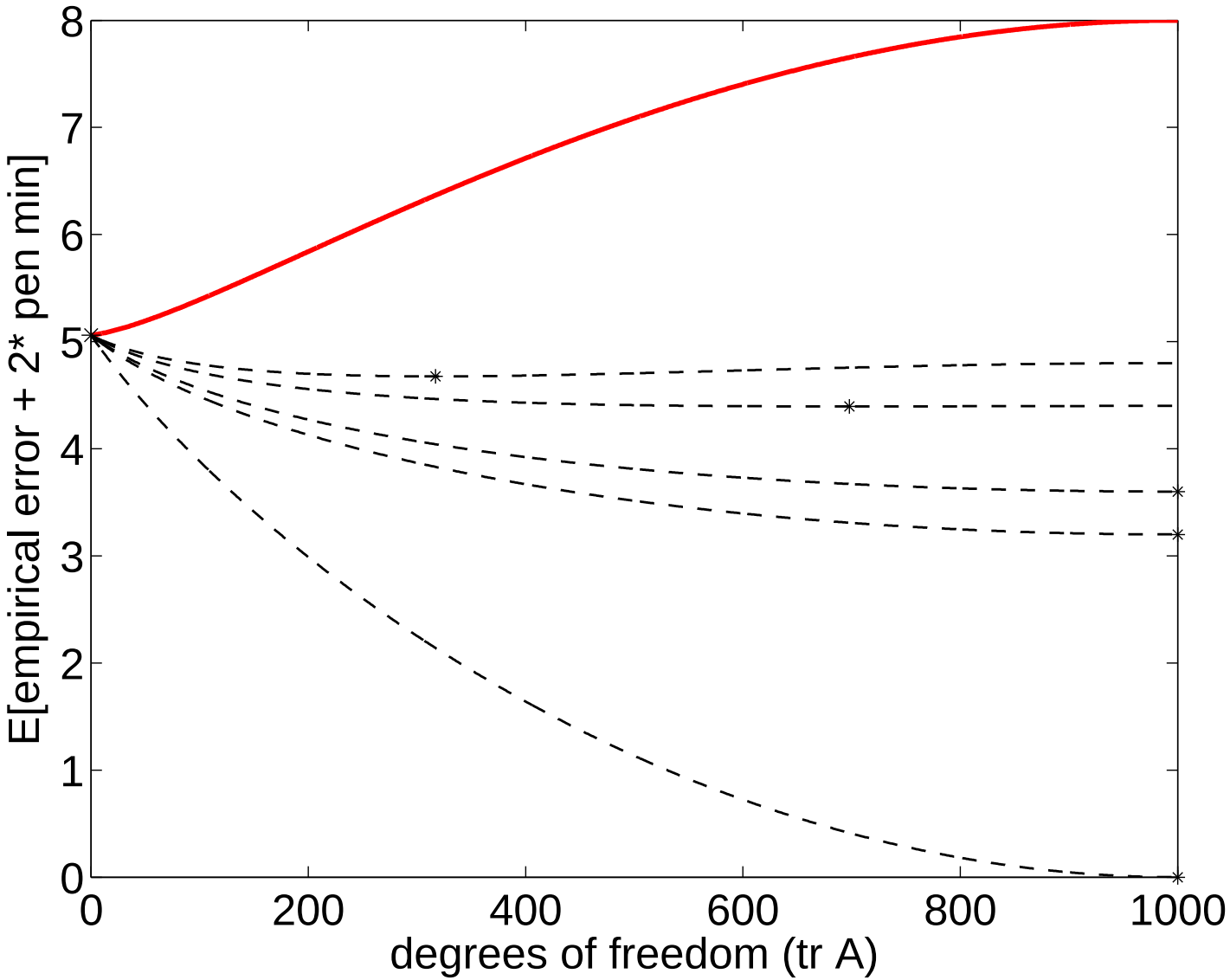
$\mathbb{E}[n^{-1}\|Y - \widehat{F}_m\|^2] + 1.1 \times \sigma^2(2\,\mathrm{tr}(A_m) - \mathrm{tr}(A_m^\top A_m))n^{-1}$



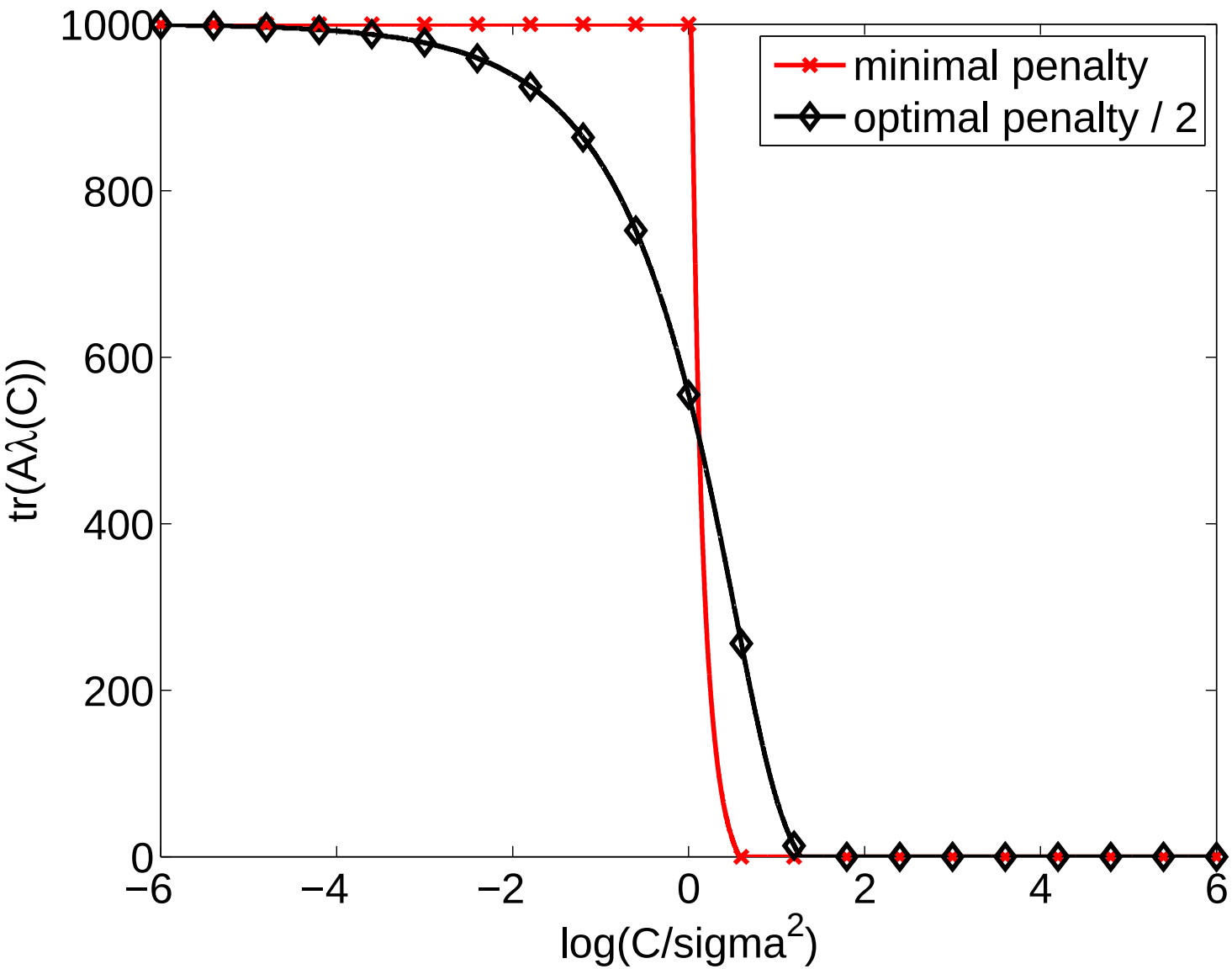
$\mathbb{E}[n^{-1}\|Y - \widehat{F}_m\|^2] + 1.2 \times \sigma^2(2\,\mathrm{tr}(A_m) - \mathrm{tr}(A_m^\top A_m))n^{-1}$



$$\mathbb{E}[n^{-1}\|Y - \widehat{F}_m\|^2] + 2 \times \sigma^2 (2\,\mathrm{tr}(A_m) - \mathrm{tr}(A_m^\top A_m)) n^{-1}$$



Saut de dimension (régression ridge)



Algorithme de calibration de pénalités (A. & Bach 2009)

1. pour tout $C > 0$, calculer

$$\widehat{m}_{\min}(C) \in \arg\min_{m \in \mathcal{M}_n} \left\{ \frac{1}{n} \left\| Y - \widehat{F}_m \right\|^2 + \frac{C \left(2 \operatorname{tr}(A_m) - \operatorname{tr}(A_m^\top A_m) \right)}{n} \right\}$$

2. trouver $\widehat{C}_{\min}$ tel que $\operatorname{tr}(A_{\widehat{m}_{\min}(C)})$ est "très grande" lorsque $C < \widehat{C}_{\min}$ et "raisonnablement petite" lorsque $C > \widehat{C}_{\min}$

3. choisir

$$\widehat{m} \in \arg\min_{m \in \mathcal{M}_n} \left\{ \frac{1}{n} \left\| Y - \widehat{F}_m \right\|^2 + \frac{2 \widehat{C}_{\min} \operatorname{tr}(A_m)}{n} \right\}$$

Comparaison avec le cas OLS

- Estimateurs linéaires :

$$\mathrm{pen}_{\mathrm{min}}(m) = \frac{\sigma^2 \left(2\,\mathrm{tr}(A_m) - \mathrm{tr}(A_m^\top A_m) \right)}{n}$$

$$\mathrm{pen}_{\mathrm{opt}}(m) = \frac{\sigma^2 \left(2\,\mathrm{tr}(A_m) \right)}{n}$$

$$\frac{\mathrm{pen}_{\mathrm{opt}}(m)}{\mathrm{pen}_{\mathrm{min}}(m)} = \frac{2\,\mathrm{tr}(A_m)}{2\,\mathrm{tr}(A_m) - \mathrm{tr}(A_m^\top A_m)} \in (1,\, 2]$$

Comparaison avec le cas OLS

- Estimateurs linéaires :

$$\mathrm{pen}_{\mathrm{min}}(m) = \frac{\sigma^2 \left(2\,\mathrm{tr}(A_m) - \mathrm{tr}(A_m^\top A_m) \right)}{n}$$

$$\mathrm{pen}_{\mathrm{opt}}(m) = \frac{\sigma^2 \left(2\,\mathrm{tr}(A_m) \right)}{n}$$

$$\frac{\mathrm{pen}_{\mathrm{opt}}(m)}{\mathrm{pen}_{\mathrm{min}}(m)} = \frac{2\,\mathrm{tr}(A_m)}{2\,\mathrm{tr}(A_m) - \mathrm{tr}(A_m^\top A_m)} \in (1, 2]$$

- Cas des moindres carrés :

$$A_m^\top A_m = A_m \quad \Rightarrow \quad \frac{\mathrm{pen}_{\mathrm{opt}}(m)}{\mathrm{pen}_{\mathrm{min}}(m)} = 2 \quad \Rightarrow \quad \text{Heuristique de pente}$$

Le cas des k plus proches voisins

$$\forall i, j \in \{1, \ldots, n\} \ , \quad A_{i,j} \in \left\{ 0, \frac{1}{k} \right\}$$

$$\forall i \in \{1, \ldots, n\} \ , \quad A_{i,i} = \frac{1}{k} \quad \text{et} \quad \sum_{j=1}^{n} A_{i,j} = 1$$

Le cas des k plus proches voisins

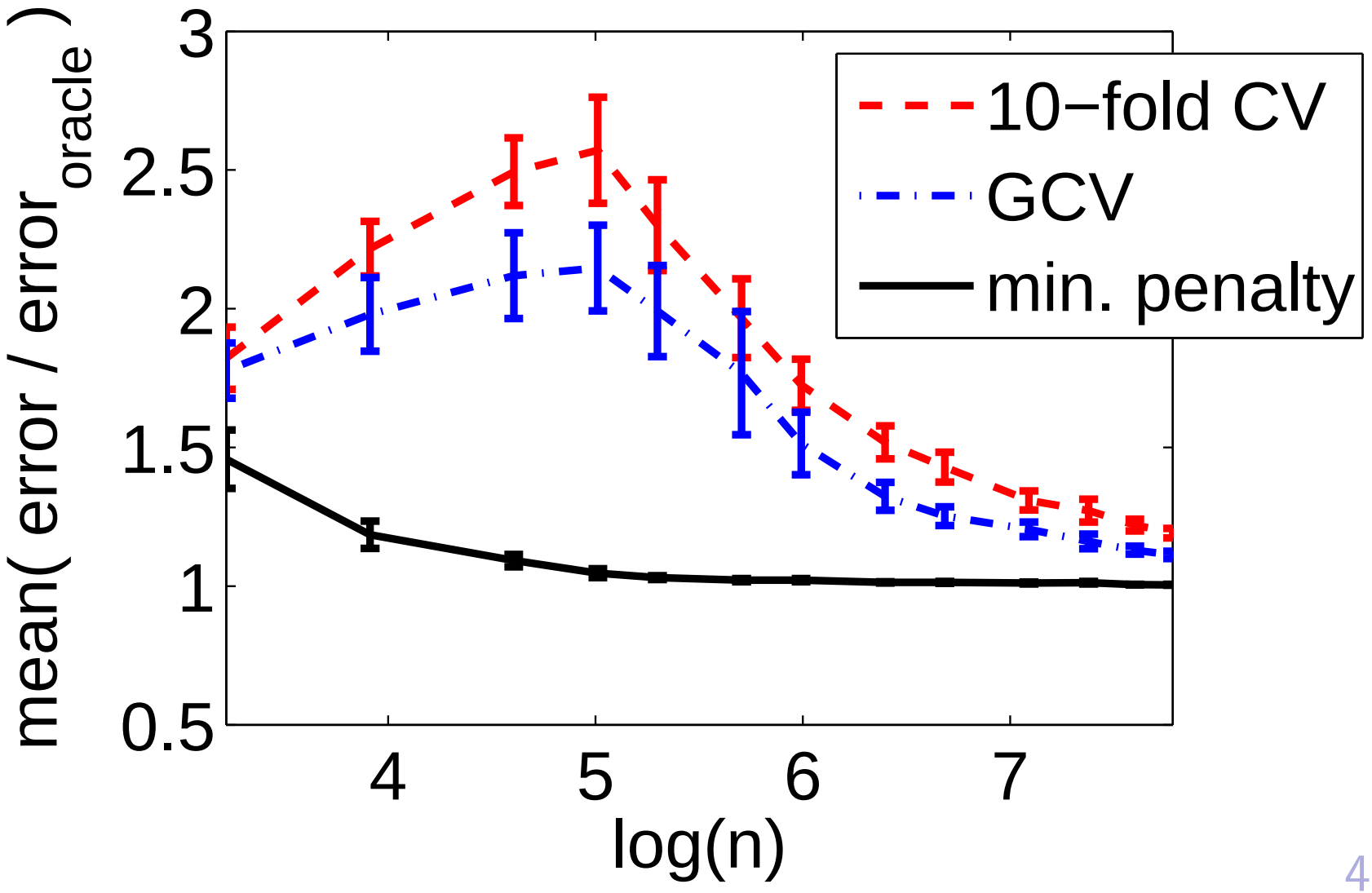
$$\forall i, j \in \{1, \ldots, n\} \ , \quad A_{i,j} \in \left\{ 0, \frac{1}{k} \right\}$$

$$\forall i \in \{1, \ldots, n\} \ , \quad A_{i,i} = \frac{1}{k} \quad \text{et} \quad \sum_{j=1}^{n} A_{i,j} = 1$$

$$\Rightarrow \quad \mathrm{tr}(A) = \frac{k}{n} = \mathrm{tr}(A^\top A)$$

$$\Rightarrow \quad \mathrm{pen}_{\mathrm{opt}} = 2\,\mathrm{pen}_{\mathrm{min}}$$

Résultats de simulation (régression ridge, choix de λ)



Théorème (1) : saut de dimension (A. & Bach, 2009)

- complexité polynomiale : $\mathrm{Card}(\mathcal{M}_n) \leq C_{\mathcal{M}} n^{\alpha}$
- bruit Gaussien homoscédastique, design déterministe
- $\exists m_1,\, m_2 \in \mathcal{M}_n$ t.q. $\mathrm{tr}(A_{m_1}) \geq n/2$, $\mathrm{tr}(A_{m_2}) \leq \sqrt{n}$ et $\forall i \in \{1,2\},\, n^{-1} \left\| (I - A_{m_i}) F \right\|^2 \leq \sigma^2 \sqrt{\ln(n)/n}$

Théorème (Pénalité minimale ; A. & Bach 2009)

Avec probabilité au moins $1 - 8 C_{\mathcal{M}} n^{-2}$, *si* $n \geq n_0(\alpha)$,

$$\forall C < \left(1 - L_\alpha \sqrt{\frac{\ln(n)}{n}} \right) \sigma^2 , \quad \mathrm{tr}\left(A_{\widehat{m}_{\min}(C)} \right) \geq \frac{n}{3}$$

$$\forall C > \left(1 + L_\alpha \frac{\sqrt{\ln(n)}}{n^{1/4}} \right) \sigma^2 , \quad \mathrm{tr}\left(A_{\widehat{m}_{\min}(C)} \right) \leq n^{3/4} \ .$$

Théorème (2) : inégalité-oracle (A. & Bach, 2009)

Hypothèse supplémentaire :

$$\exists \kappa \geq 1\,,\ \forall m \in \mathcal{M}_n\,, \quad \operatorname{tr}(A_m)\sigma^2 \leq \kappa \mathbb{E}\left[\left\|F - \widehat{F}_m\right\|^2\right]$$

Théorème (Inégalité-oracle ; A. & Bach 2009)

Alors, avec probabilité au moins $1 - 8C_{\mathcal{M}}n^{-2}$, *si* $n \geq n_0(\alpha)$,

$$\forall \widehat{m} \in \arg\min_{m \in \mathcal{M}} \left\{ \frac{1}{n}\left\|Y - \widehat{F}_m\right\|^2 + \frac{2\widehat{C}\operatorname{tr}(A_m)}{n} \right\}\,,$$

$$\frac{1}{n}\left\|F - \widehat{F}_{\widehat{m}}\right\|^2 \leq \left(1 + \frac{40\kappa}{\ln(n)}\right) \inf_{m \in \mathcal{M}_n} \left\{ \frac{1}{n}\left\|F - \widehat{F}_m\right\|^2 \right\} + \frac{36(\kappa + \alpha + 2)\ln(n)\sigma^2}{n}\,.$$

Plan

Pénalités minimales : cas général

$$\widehat{m}_{\text{min}}(C) \in \arg\min_{m \in \mathcal{M}_n} \left\{ P_n \gamma\left(\widehat{s}_m\right) + C\, \text{pen}_0(m) \right\}$$

- explosion du risque pour $C < C^\star_{\text{min}}$
- inégalité oracle pour $C > C^\star_{\text{min}}$

Pénalités minimales : cas général

$$\widehat{m}_{\text{min}}(C) \in \arg\min_{m \in \mathcal{M}_n} \left\{ P_n \gamma\left(\widehat{s}_m\right) + C\, \text{pen}_0(m) \right\}$$

- explosion du risque pour $C < C^\star_{\text{min}}$
- inégalité oracle pour $C > C^\star_{\text{min}}$

- saut de complexité $\mathcal{C}_{\widehat{m}(C)}$ autour de $C = C^\star_{\text{min}}$
- $\text{pen}_{\text{min}} = C^\star_{\text{min}}\, \text{pen}_0$ avec pen_0 connue (estimable)

Pénalités minimales : cas général

$$\widehat{m}_{\text{min}}(C) \in \arg\min_{m \in \mathcal{M}_n} \left\{ P_n \gamma\left(\widehat{s}_m\right) + C\, \text{pen}_0(m) \right\}$$

- explosion du risque pour $C < C^\star_{\text{min}}$
- inégalité oracle pour $C > C^\star_{\text{min}}$
- saut de complexité $\mathcal{C}_{\widehat{m}(C)}$ autour de $C = C^\star_{\text{min}}$
- $\text{pen}_{\text{min}} = C^\star_{\text{min}}\, \text{pen}_0$ avec pen_0 connue (estimable)
- inégalité-oracle optimale lorsque $\text{pen} = \text{pen}_{\text{opt}}$
- pen_{opt} connue (estimable) à $C^\star$ près

Pénalités minimales : cas général

$$\widehat{m}_{\text{min}}(C) \in \arg\min_{m \in \mathcal{M}_n} \left\{ P_n \gamma\left(\widehat{s}_m\right) + C \, \text{pen}_0(m) \right\}$$

- explosion du risque pour $C < C^\star_{\text{min}}$
- inégalité oracle pour $C > C^\star_{\text{min}}$
- saut de complexité $\mathcal{C}_{\widehat{m}(C)}$ autour de $C = C^\star_{\text{min}}$
- $\text{pen}_{\text{min}} = C^\star_{\text{min}} \, \text{pen}_0$ avec pen_0 connue (estimable)
- inégalité-oracle optimale lorsque $\text{pen} = \text{pen}_{\text{opt}}$
- pen_{opt} connue (estimable) à $C^\star$ près
- lien connu entre $C^\star_{\text{min}}$ et $C^\star$

Calibration à l'aide de pénalités minimales

Données : $\mathrm{pen}_0 = \dfrac{1}{C^\star_{\min}} \mathrm{pen}_{\min}$ $\mathrm{pen}_1 = \dfrac{1}{f(C^\star_{\min})} \mathrm{pen}_{\mathrm{opt}}$ f

1. pour tout $C > 0$, calculer

$$\widehat{m}_{\min}(C) \in \arg\min_{m \in \mathcal{M}_n} \left\{ P_n \gamma\left(\widehat{s}_m\right) + C\, \mathrm{pen}_0(m) \right\} \quad ,$$

2. trouver $\widehat{C}_{\min}$ tel que $C_{\widehat{m}_{\min}(C)}$ est "très grande" lorsque $C < \widehat{C}_{\min}$ et "raisonnablement petite" lorsque $C > \widehat{C}_{\min}$,

3. choisir

$$\widehat{m} \in \arg\min_{m \in \mathcal{M}_n} \left\{ P_n \gamma\left(\widehat{s}_m\right) + f\left(\widehat{C}_{\min}\right) \mathrm{pen}_1(m) \right\} \quad ,$$

Résultats mathématiques existants

- Concentration de p_2 : minimum de contraste borné (Boucheron & Massart, 2010)
- Explosion du risque si $C < C^\star_{\text{min}}$:
  - OLS, bruit Gaussien, complexité de $\mathcal{M}_n$ exponentielle (Birgé & Massart, 2007)
  - OLS, bruit Gaussien, complexité de $\mathcal{M}_n$ intermédiaire et $s^\star = 0$ (Birgé & Massart, 2007)
  - Densité, estimateur de Dantzig, dans un cas particulier (Bertin, Le Pennec et Rivoirard, 2009)
  - Estimation de l'intensité d'un processus de Poisson par seuillage (Reynaud-Bouret et Rivoirard, 2009), avec $C^\star / C^\star_{\text{min}} \in [1, 12]$
- Explosion de la dimension si $C < C^\star_{\text{min}}$: OLS, bruit Gaussien, pénalités multiplicatives, complexité de $\mathcal{M}_n$ polynômiale ou exponentielle et $s^\star = 0$ (Baraud, Giraud et Huet, 2009)

Plan

Avantages pratiques de la méthode

- vérification visuelle de l'existence d'un saut
- calibration indépendante du choix d'un modèle m_0
- sur-apprentissage trop fort pratiquement impossible
- un paramètre à régler : la manière de localiser le saut

Comment localiser le saut effectivement ?

- Saut de dimension : plus grand saut ? plus grand saut relatif ? valeur seuil de la dimension ?

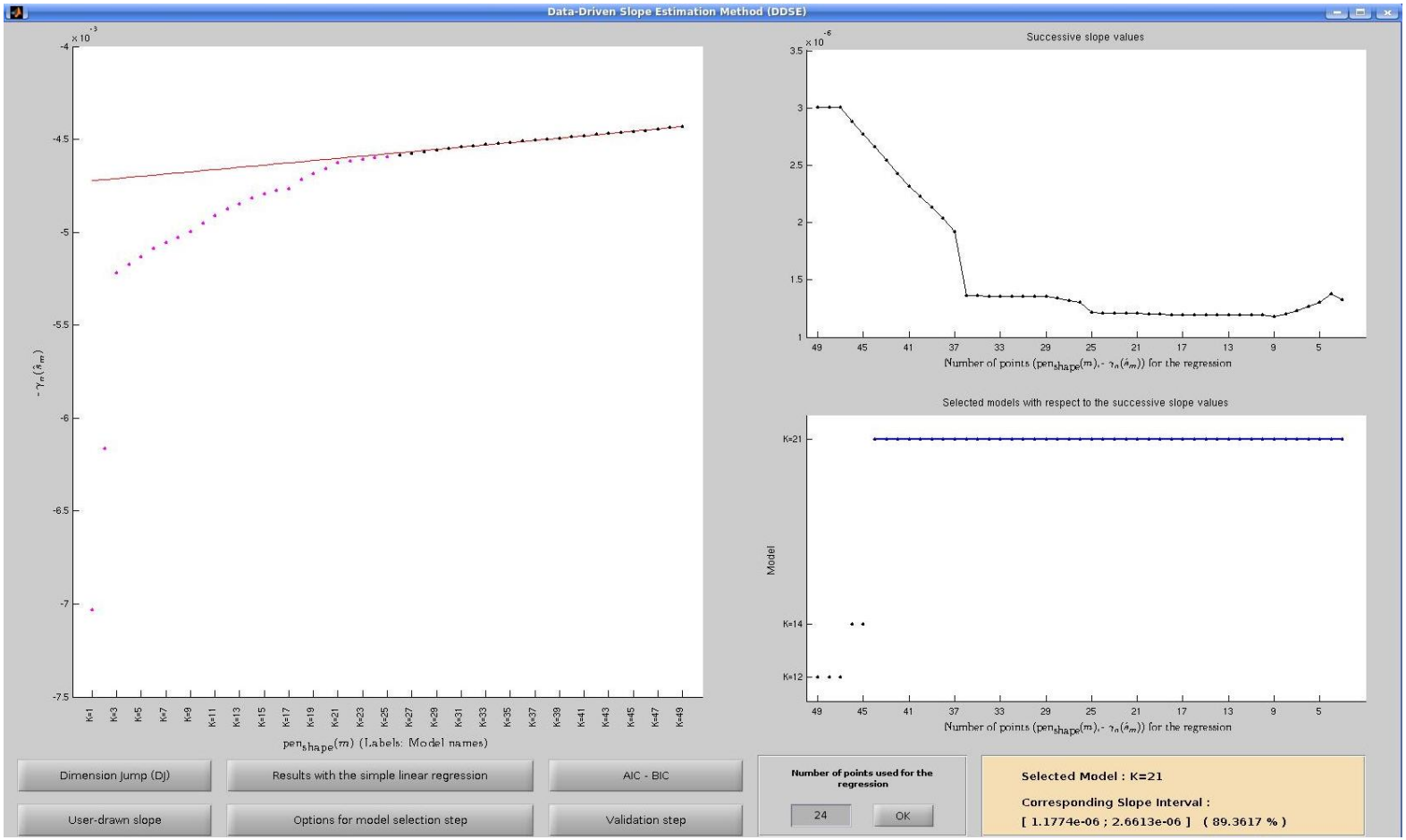
Comment localiser le saut effectivement ?

- Saut de dimension : plus grand saut ? plus grand saut relatif ? valeur seuil de la dimension ?
- Estimation de la pente du risque empirique : avec quels modèles la calculer ? régression robuste ?

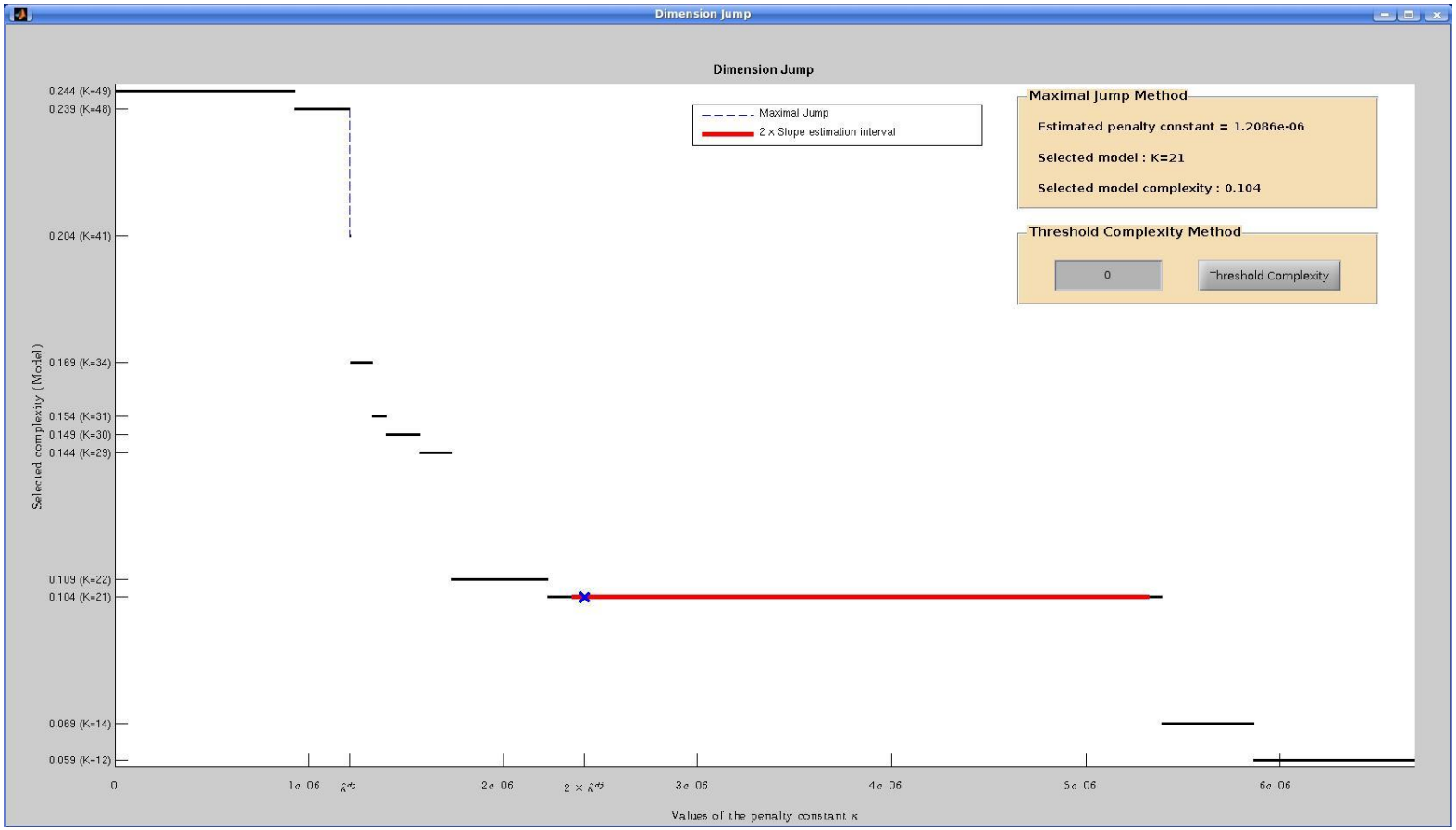
Comment localiser le saut effectivement ?

- Saut de dimension : plus grand saut ? plus grand saut relatif ? valeur seuil de la dimension ?
- Estimation de la pente du risque empirique : avec quels modèles la calculer ? régression robuste ?
- Saut ou pente ? Les deux !
  $\Rightarrow$ package CAPUSHE (Baudry, Maugis & Michel, 2010)
  `http://www.math.univ-toulouse.fr/~maugis/CAPUSHE.html`

CAPUSHE (Baudry, Maugis & Michel, 2010) : pente

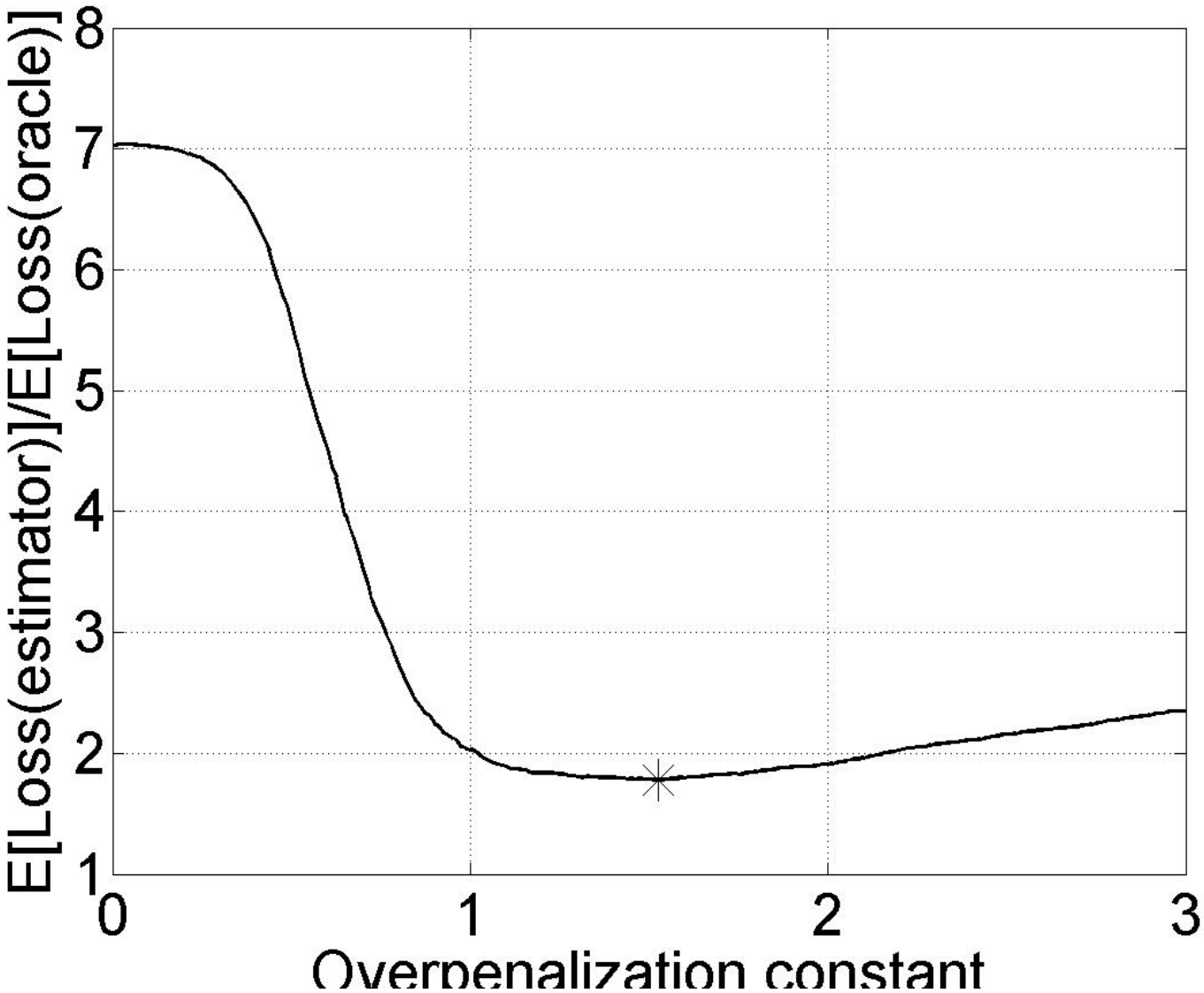


CAPUSHE (Baudry, Maugis & Michel, 2010) : saut



Plan

Surpénalisation



Plan des séances restantes

- Lundi 24, 14h–16h : Rééchantillonnage et pénalisation
- Lundi 31, 15h30–17h30 : Validation croisée et pénalités reliées