arxiv: 1409.6496
A. + P. Mathé.

$$\boxed{z = Tx + \delta\xi}$$

- $x \in X$ unknown
- z obs
- $T : X \to Y$ bdd op (blurring)
- $\xi \sim N(0, I)$ noise, $\delta > 0$ known noise level.

Prior: $x \sim N(m_0, \frac{\delta^2}{\alpha} C_0)$, $\alpha > 0$ scaling parameter (exact meaning in a moment)

C_0: > 0, t-c
assume C_0, T commuting

Likelihood: $z | x \sim N(Tx, \delta^2 I)$ (linear model density of prior & likelihood quadratic in x)

$\Rightarrow$ Posterior: $x | z \sim N(x_\alpha^\delta, C(\alpha, \delta)) = \mu_{\alpha,\delta}^z$

1) Introduce Setup
2) Posterior Contraction
3) Delaying saturation by preconditioning the prior
(4) General technique for getting ROC)

Complete the square

$$C^{-1} = \frac{\alpha}{\delta^2} C_0^{-1} + \frac{1}{\delta^2} T^* T = \delta^{-2} C_0^{-1/2} (\alpha I + B^* B) C_0^{-1/2}$$

$B = T C_0^{1/2}$

compact

$$C^{-1}(x_\alpha^\delta - m_0) = \delta^{-2} T^* (z - T m_0)$$

$$\Leftrightarrow \quad \boxed{x_\alpha^\delta = C_0^{1/2} (\alpha I + B^* B)^{-1} B^* (z - T m_0)} + m_0$$

ignore Tm_0, x_α^δ is regularized inversion of data,

α acts as reg. parameter

(if $\alpha = 0$ " $x_0^\delta = C_0^{1/2} (B^* B)^{-1} B^* z = T^{-1} z$)

__Posterior Consistency__: Assume $z = z^\delta = T x^* + \delta \xi$

small noise limit

x^* fixed truth

$$\boxed{\text{As } \delta \to 0, \text{ can we choose } \alpha = \alpha(\delta) \to 0 \text{ s.t. } \text{"} \mu_{\alpha,\delta}^{z=z^\delta} \to \delta_{x^*} \text{" ?}}$$

reg needs to disappear

For us "$\to$" is in the sense of SPC

$$SPC = \mathbb{E}^{X^*} \mathbb{E}_{\alpha}^{z^\delta} \| x^* - x \|^2$$

wrt data generating measure $N(Tx^*, \delta I)$.

wrt posterior $\mu_{\alpha,\delta}^{z^\delta}$

$$= \mathbb{E}^{X^*} \left(\| x^* - x_\alpha^\delta \|^2 + \operatorname{tr}(C(\alpha,\delta)) \right)$$

bias-variance decomposition of ~~...~~ expectation wrt post

indep of data

$$= \| x^* - \mathbb{E}^{X^*} x_\alpha^\delta \|^2 + \mathbb{E}^{X^*} \| x_\alpha^\delta - \mathbb{E}^{X^*} x_\alpha^\delta \|^2 + \operatorname{tr}(C(\alpha,\delta))$$

bias-variance of ~~...~~ expectation wrt data

squared bias (of post mean as estimator of X^*)

estimation variance

posterior spread

Assume $x^* \in X^\gamma$, γ quantifies regularity assumed to be known (unrealistic)

As $\delta \to 0$, choose $\alpha = \alpha(\delta, \gamma) \to 0$ s.t. $SPC \to 0$

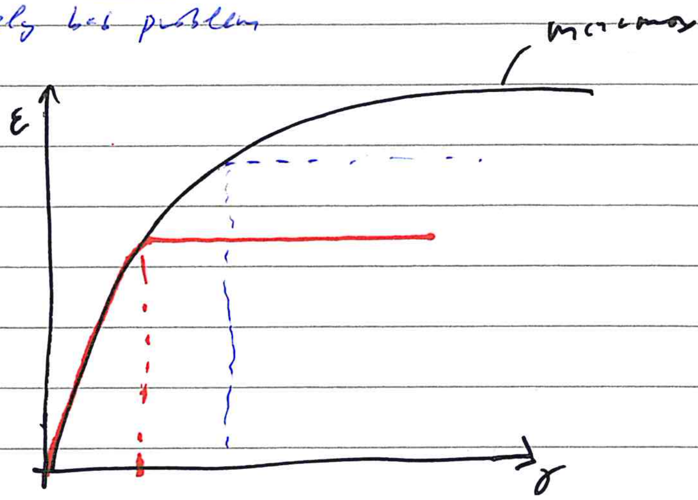
at optimal rate (minimax over ~~...~~ X^γ).

eg KVZ11, ALS13, KVZ12, ASZ14

moderately bad problem — severely bad problem

Typical picture

$SPC \simeq \delta^{\varepsilon}$



saturation phenomenon

due to saturation of bias

(since only term which depends on x^*)

Contributions: a) Use classic reg. tools to get SPC rates in general settings (before relied on explicit calcs)

b) Show that choosing prior mean as regularised inverse of data delays or even removes saturation in bias (hence SPC)

will try to explain this. If I have time I will give simplified version of a)

focus on bias

If $m_0 = 0$ bias $b_{x^*}(\alpha) = \| x^* - \mathbb{E}^{x^*} x_\alpha^\delta \|$

$$= \| x^* - C_0^{1/2} (\alpha I + B^* B)^{-1} B^* T x^* \|$$

$$= \| (I - (\alpha I + B^* B)^{-1} B^* B) x^* \|$$

$$= \alpha \underbrace{\| (\alpha I + B^* B)^{-1} x^* \|}_{\text{at best } O(1) \text{ as } \alpha \to 0}$$

This happens if $x^* \in R(B^* B)$
so that $(B^* B)^{-1} x^* \in X$

If not, norm blow up as $\alpha \to 0$, rate slower than $1/\alpha$ (provided $x^* \in X$) (so that $\alpha \cdot \|\cdot\| \to 0$)

Choose $\boxed{m_0 = m_\alpha^\delta = C_0^{1/2} g_\alpha (B^* B) B^* z^\delta}$

reg. filter

- $g_\alpha(B^* B)$ bdd $\forall \alpha > 0$
- $g_\alpha(B^* B) \xrightarrow{\alpha \to 0} (B^* B)^{-1}$
- associated residual fcn $r_\alpha(t) = 1 - t g_\alpha$.
  $r_\alpha \xrightarrow{\alpha \to 0} 0$.

~~eg Tikhonov:~~ $\cancel{g_\alpha(t) = \frac{1}{\alpha + t}}$

$\cancel{r_\alpha(t) = \frac{\alpha}{\alpha + t}}$

~~Spectral cut-off:~~ $\cancel{g_\alpha(t) = \begin{cases} \frac{1}{t}, & t \geq \alpha \\ 0, & t < \alpha \end{cases}}$

$\cancel{r_\alpha(t) = \begin{cases} 0, & t \geq \alpha \\ 1, & t < \alpha \end{cases}}$

eg **Tikhonov:** $g_\alpha(t) = \frac{1}{\alpha + t}$, $r_\alpha(t) = \frac{\alpha}{\alpha + t}$ (very induced by Lasso prior for $m \neq 0$)

Spectral cut-off: $g_\alpha(t) = \begin{cases} 1/t, & t \geq \alpha \\ 0, & t < \alpha \end{cases}$

$r_\alpha(t) = \begin{cases} 0, & t \geq \alpha \\ 1, & t < \alpha \end{cases}$

Then $b_{x^*}(\alpha) = \cdots = \alpha \, \| (\alpha + B^*B)^{-1} r_\alpha(B^*B) x^* \|$

eg g_α **Tikhonov** $= \alpha^2 \, \underbrace{\| (\alpha + B^*B)^{-2} x^* \|}_{\text{can at best be } O(1)}$

for $x^* \in R((B^*B)^2)$

If not, norm blows up slower than $1/\alpha^2$

Delayed saturation! (full picture)

Now try to explain α)

- φ <u>index fcn</u> if cts nondecreasing, $\varphi(0)=0$.
- <u>Qualification</u>: φ qualification for reg g_α if

$$r_\alpha(t)\varphi(t) \le C\varphi(\alpha) \qquad \forall \alpha, t.$$

quantifies ability of reg to take smoothness into account.

eg Tikhonov $r_\alpha(t) = \frac{\alpha}{\alpha+t}\, t \le \alpha$ (this is what we have for $\underline{m_0=0}$)

so that $\varphi(t)=t$ (maximal) qualification.

- If φ qualification for g_α, then any $\psi \prec \varphi$ also qualification for g_α.

goes to 0 slower than t

eg $\frac{\alpha}{\alpha+t}\, t^{1/2} \le \alpha^{1/2}$ since $\alpha^{1/2} t^{1/2} \le \alpha + t$.

eg spectral cut off $r_\alpha(t) = \begin{cases} 0, & t \ge \alpha \\ 1, & t < \alpha \end{cases}$

$r_\alpha(t)\varphi(t) = 0 \le \varphi(\alpha) \quad t \ge \alpha$

$r_\alpha(t)\varphi(t) = \varphi(t) \le \varphi(\alpha), \quad t \le \alpha$ (φ non decreasing)

TSVD has arbitrary qualification.

Assume $x^* \in A_\varphi = \{x : x = \varphi(B^*B)w, \quad \|w\| \leq 1\}$

Proposition:

i) If $\varphi(t) \prec t$, $\quad b_{x^*}(\alpha) \leq c\,\varphi(\alpha)$

(low smoothness)

independently of m_0.

ii) If $t \prec \varphi(t)$ and $m_0 = 0$, $\quad b_{x^*}(\alpha) \asymp \alpha$

(high smoothness)

iii) If $t \prec \varphi(t)$ and $\dfrac{\varphi(t)}{t}$ qualification for g_α

$$b_{x^*}(\alpha) \leq c\,\varphi(\alpha).$$

Proof:

i) $\|\alpha(\alpha I + B^*B)^{-1}\varphi(B^*B)w\| \leq \|\alpha(\alpha I + B^*B)^{-1}\varphi(B^*B)\|$

$\leq \varphi(\alpha)$

Tikhonov qualification is t, $\varphi \prec t$

ii) discussed earlier

iii) $\|\alpha(\alpha I + B^*B)^{-1}\, r_\alpha(B^*B)\,\varphi(B^*B)\|$

$= \|\alpha(\alpha I + B^*B)^{-1}(B^*B)\,(B^*B)^{-1}\, r_\alpha(B^*B)\,\varphi(B^*B)\|$

$\leq \|\alpha(\alpha I + B^*B)^{-1}B^*B\|\;\|r_\alpha(B^*B)\,\varphi(B^*B)(B^*B)^{-1}\|$

$$\underset{\downarrow\text{Tikhonov}}{\leq}\; \alpha \; \frac{\varphi(\alpha)}{\alpha} = \varphi(\alpha)$$

eg g_α Tikhonov. Then for $\varphi(t) = t^2$, $\dfrac{\varphi(t)}{t} = t$ qualification for g_α

$\Rightarrow b_{x^*}(\alpha) \leq \alpha^2$ (delayed saturation)

eg g_α spectral cut-off. never saturates

To allow for more flexibility define

$$A_\varphi = \{x : x = \varphi(C_0)w, \ \|w\| \leq 1\}$$

and introduce link conditions between T and C_0.