Maximum a posteriori testing in statistical inverse problems

Remo Kretschmann

Seminar on Computational Engineering, LUT
Lappeenranta, 20 November 2024

Joint work with

Frank Werner and Daniel Wachsmuth

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Structure

Set-up

Consider statistical linear inverse problem

$$Y = Tu^\dagger + \sigma Z,$$

where

- $T\colon \mathcal{X} \to \mathcal{Y}$ bounded linear forward operator between real separable Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$,
- $u^\dagger \in \mathcal{X}$ unknown quantity of interest,
- $\sigma > 0$ noise level,
- Z white Gaussian noise process on $\mathcal{Y}$.

For each $g \in \mathcal{Y}$ one has access to real-valued Gaussian random variable

$$\langle Y, g \rangle = \left\langle Tu^\dagger, g \right\rangle_{\mathcal{Y}} + \sigma \langle Z, g \rangle .$$

Feature inference

- $\mathcal{X}$,$\mathcal{Y}$ typically function spaces such as $L^p(\Omega)$ or $H^s(\Omega)$ on some domain $\Omega \subseteq \mathbb{R}^d$.
- Often one is not interested in whole function $u^\dagger$ but in certain features of it such as modes, homogeneity, monotonicity, or support.
- Many features can be described by (family of) bounded linear functionals $\varphi \in \mathcal{X}^*$.
- We perform inference for such features by means of statistical hypothesis testing. Specifically, we test

$$H_0 : \left\langle \varphi, u^\dagger \right\rangle_{\mathcal{X}^* \times \mathcal{X}} \leq 0 \quad \text{against} \quad H_1 : \left\langle \varphi, u^\dagger \right\rangle_{\mathcal{X}^* \times \mathcal{X}} > 0.$$

Example 1: Support inference in deconvolution

- T convolution operator
$$Tu = h * u$$
on $L^2(\mathbb{R})$ with kernel h.
- Question: Is $\operatorname{supp} u^\dagger \cap (a, b) = \emptyset$?
- Under assumption that $u^\dagger$ is nonnegative, indicator function $\varphi := \mathbf{1}_{[a,b]}$ describes feature of interest
$$\left\langle \varphi, u^\dagger \right\rangle_{L^2} = \int_a^b u^\dagger(x) \mathrm{d}x.$$

Example 2: Linearity inference

- Direct noisy measurement
$$Y = f^\dagger + \sigma Z$$
of function $f^\dagger \in H_0^1(0,1) \cap H^2(0,1)$.
- Question: Is $f^\dagger$ linear on $(a,b) \subseteq (0,1)$?
- For $u \in L^2(0,1)$, let $Tu = f$ be weak solution to
$$-f'' = u \quad \text{on } (0,1), \qquad f(0) = f(1) = 0.$$
- Under assumption that $f^\dagger$ is concave, $\varphi := \mathbf{1}_{[a,b]}$ describes feature of interest
$$\left\langle \varphi, u^\dagger \right\rangle_{L^2} = -\int_a^b (f^\dagger)''(x) \mathrm{d}x.$$

Basic properties of hypothesis tests

- Hypothesis test $\Psi(Y)$ takes only values 0 (accepts) and 1 (rejects).
- Probability of rejection $\mathbb{P}_{u^\dagger}[\Psi(Y) = 1]$ is called size of test Ψ for $u^\dagger$.

Basic properties of hypothesis tests

- Hypothesis test $\Psi(Y)$ takes only values 0 (accepts) and 1 (rejects).
- Probability of rejection $\mathbb{P}_{u^\dagger}[\Psi(Y)=1]$ is called size of test Ψ for $u^\dagger$.

Probability of false rejection

Maximal size of test under hypothesis H_0

$$\sup\left\{\mathbb{P}_{u^\dagger}[\Psi(Y)=1] : u^\dagger \in \mathcal{X} \text{ satisfies } H_0\right\}$$

is called level (of significance) of test Ψ.

Basic properties of hypothesis tests

- Hypothesis test $\Psi(Y)$ takes only values 0 (accepts) and 1 (rejects).
- Probability of rejection $\mathbb{P}_{u^\dagger}[\Psi(Y) = 1]$ is called size of test Ψ for $u^\dagger$.

Probability of false rejection

Maximal size of test under hypothesis H_0

$$\sup \left\{ \mathbb{P}_{u^\dagger}[\Psi(Y) = 1] : u^\dagger \in \mathcal{X} \text{ satisfies } H_0 \right\}$$

is called level (of significance) of test Ψ.

Probability of correct rejection

Size of test under alternative H_1 is also called power of test Ψ for $u^\dagger$.

Unregularized hypothesis testing[1]

- Assume that $\varphi \in \operatorname{ran} T^*$ and choose $\Phi_0 \in \mathcal{Y}$ such that

$$T^*\Phi_0 = \varphi.$$

- Then $\langle Y, \Phi_0 \rangle$ is natural estimator for desired quantity

$$\langle \varphi, u^\dagger \rangle_{\mathcal{X}} = \langle T^*\Phi_0, u^\dagger \rangle_{\mathcal{X}} = \langle \Phi_0, T u^\dagger \rangle_{\mathcal{Y}}.$$

- Define test

$$\Psi_0(Y) := \mathbf{1}_{\langle Y, \Phi_0 \rangle > c}, \quad c \in \mathbb{R}.$$

[1] K. Proksch, F. Werner, A. Munk (2018). *Multiscale scanning in inverse problems*. Ann. Statist., 46(6B).

Unregularized hypothesis testing[1]

- Assume that $\varphi \in \operatorname{ran} T^*$ and choose $\Phi_0 \in \mathcal{Y}$ such that

$$T^*\Phi_0 = \varphi.$$

- Then $\langle Y, \Phi_0 \rangle$ is natural estimator for desired quantity

$$\langle \varphi, u^\dagger \rangle_{\mathcal{X}} = \langle T^*\Phi_0, u^\dagger \rangle_{\mathcal{X}} = \langle \Phi_0, Tu^\dagger \rangle_{\mathcal{Y}}.$$

- Define test

$$\Psi_0(Y) := \mathbf{1}_{\langle Y, \Phi_0 \rangle > c}, \quad c \in \mathbb{R}.$$

- For any $\alpha \in (0,1)$, critical value $c = c(\varphi, \alpha)$ can be chosen such that test Ψ_0 has level α for testing H_0 against H_1.

[1] K. Proksch, F. Werner, A. Munk (2018). *Multiscale scanning in inverse problems*. Ann. Statist., 46(6B).

Issues

- Unregularized level α test has power

$$\mathbb{P}_{u^\dagger}\left[\Psi_0(Y) = 1\right] = Q\left(Q^{-1}(\alpha) + \frac{\langle \varphi, u^\dagger \rangle}{\sigma \left\|\Phi_0\right\|}\right).$$

Issues

- Unregularized level α test has power

$$\mathbb{P}_{u^\dagger}\left[\Psi_0(Y) = 1\right] = Q\left(Q^{-1}(\alpha) + \frac{\langle \varphi, u^\dagger \rangle}{\sigma \left\| \Phi_0 \right\|}\right).$$

- For certain features, unregularized testing is unfeasable.
  1. If $\varphi \notin \operatorname{ran} T^*$, approach not applicable.
  2. Probe element Φ_0 is solution to ill-posed equation

$$T^*\Phi_0 = \varphi.$$

  For certain features, norm of Φ_0 is huge, and power of unregularized test Ψ_0 is arbitrarily close to level.

Solutions

Both of these limitations can be overcome by regularized hypothesis tests

$$\Psi_{\Phi,c}(Y) := \mathbf{1}_{\langle Y,\Phi\rangle > c}, \quad \Phi \in \mathcal{Y}, c \in \mathbb{R}.$$

1. Maximize (empirical) power among class of regularized level α tests[2].
2. Define tests using Bayesian approach: Reject based upon posterior probabilities.
3. Choose probe element Φ as Tikhonov regularized solution to equation $T^*\Phi_0 = \varphi$.

[2] R. Kretschmann, D. Wachsmuth, F. Werner (2024). *Optimal regularized hypothesis testing in statistical inverse problems*. Inverse Problems 40, 015013.

Questions

1. Do Bayesian tests fall under the framework of regularized hypothesis testing?
2. How do they relate to Tikhonov regularized tests?
3. Do they overcome aforementioned issues? Do they relieve the restrictions on φ?
4. Do they have a high power?

Structure

Bayesian set-up

Consider problem from Bayesian perspective,

$$Y = TU + \sigma Z.$$

- Assign Gaussian prior distribution $\Pi = \mathcal{N}(m_0, C_0)$ to U,
- C_0 symmetric, positive definite, trace class,
- U and Z independent.

Conditional distribution of U, given $Y = y$, almost surely Gaussian $\mathcal{N}(m, C)$ with

$$C = \sigma^2 C_0^{\frac{1}{2}} \left(C_0^{\frac{1}{2}} T^* T C_0^{\frac{1}{2}} + \sigma^2 \mathrm{Id} \right)^{-1} C_0^{\frac{1}{2}},$$

$$m = m_0 + C_0^{\frac{1}{2}} \left(C_0^{\frac{1}{2}} T^* T C_0^{\frac{1}{2}} + \sigma^2 \mathrm{Id} \right)^{-1} C_0^{\frac{1}{2}} T^* (y - T m_0).$$

Maximum a posteriori testing

For $\varphi \in \mathcal{X}$, define maximum a posteriori (MAP) test Ψ_{MAP} by

$$\Psi_{\mathsf{MAP}}(y) := \begin{cases} 1 & \text{if } \mathbb{P}\left[\langle \varphi, U \rangle > 0 | Y = y\right] > \mathbb{P}\left[\langle \varphi, U \rangle \leq 0 | Y = y\right], \\ 0 & \text{otherwise.} \end{cases}$$

- Study properties of Ψ_{MAP} in frequentistic setting.
- Conditional distribution of $\langle \varphi, U \rangle_{\mathcal{X}}$, given $Y = y$, is

$$\mathcal{N}\left(\langle \varphi, m \rangle_{\mathcal{X}}, \langle \varphi, C\varphi \rangle_{\mathcal{X}}\right).$$

Maximum a posteriori testing

For $\varphi \in \mathcal{X}$, define maximum a posteriori (MAP) test Ψ_{MAP} by

$$\Psi_{\text{MAP}}(y) := \begin{cases} 1 & \text{if } \mathbb{P}\left[\langle \varphi, U \rangle > 0 | Y = y\right] > \frac{1}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

- Study properties of Ψ_{MAP} in frequentistic setting.
- Conditional distribution of $\langle \varphi, U \rangle_{\mathcal{X}}$, given $Y = y$, is

$$\mathcal{N}\left(\langle \varphi, m \rangle_{\mathcal{X}}, \langle \varphi, C\varphi \rangle_{\mathcal{X}}\right).$$

Evaluating MAP test

- Cdf F_φ of $\langle \varphi, U \rangle_{\mathcal{X}}$, given $Y = y$, is

$$F_\varphi(t) = \mathbb{P}\left[\langle \varphi, U \rangle \leq t \middle| Y = y\right] = Q\left(\frac{t - \langle \varphi, m \rangle}{\langle \varphi, C\varphi \rangle^{1/2}}\right),$$

where Q is cdf of $\mathcal{N}(0,1)$.

Evaluating MAP test

- Cdf F_φ of $\langle \varphi, U \rangle_{\mathcal{X}}$, given $Y = y$, is

$$F_\varphi(t) = \mathbb{P}\left[\langle \varphi, U \rangle \leq t | Y = y\right] = Q\left(\frac{t - \langle \varphi, m \rangle}{\langle \varphi, C\varphi \rangle^{1/2}}\right),$$

  where Q is cdf of $\mathcal{N}(0,1)$.
- Hence

$$\begin{aligned} \Psi_{\mathrm{MAP}}(y) = 1 \quad &\Leftrightarrow \quad \mathbb{P}\left[\langle \varphi, U \rangle_{\mathcal{X}} > 0 | Y = y\right] > \frac{1}{2} \\ &\Leftrightarrow \quad F_\varphi(0) < \frac{1}{2} \quad \Leftrightarrow \quad \langle \varphi, m \rangle_{\mathcal{X}} > 0. \end{aligned}$$

Connection with Tikhonov regularization

- We have

$$\langle \varphi, m \rangle_{\mathcal{X}} = \langle y, \Phi_{\mathrm{MAP}} \rangle - \langle m_0, T^* \Phi_{\mathrm{MAP}} - \varphi \rangle_{\mathcal{X}} \,,$$

  where

$$\Phi_{\mathrm{MAP}} := T C_0^{\frac{1}{2}} \left(C_0^{\frac{1}{2}} T^* T C_0^{\frac{1}{2}} + \sigma^2 \mathrm{Id} \right)^{-1} C_0^{\frac{1}{2}} \varphi.$$

Connection with Tikhonov regularization

- We have
$$\langle \varphi, m \rangle_{\mathcal{X}} = \langle y, \Phi_{\mathrm{MAP}} \rangle - \langle m_0, T^* \Phi_{\mathrm{MAP}} - \varphi \rangle_{\mathcal{X}} \,,$$
where
$$\Phi_{\mathrm{MAP}} := T C_0^{\frac{1}{2}} \left(C_0^{\frac{1}{2}} T^* T C_0^{\frac{1}{2}} + \sigma^2 \mathrm{Id} \right)^{-1} C_0^{\frac{1}{2}} \varphi.$$

- If T is compact and C_0 commutes with $T^* T$, then Φ_{MAP} is minimizer of
$$\Phi \mapsto \| T^* \Phi - \varphi \|_{\mathcal{X}}^2 + \sigma^2 \left\| C_0^{-\frac{1}{2}} V^* \Phi \right\|_{\mathcal{X}}^2 ,$$
where V is a unitary operator such that $T = V |T|$.

Interpretation as regularized test

MAP test Ψ_{MAP} corresponds to regularized test $\Psi_{\Phi_{\mathrm{MAP}}, c_{\mathrm{MAP}}}$ with

$$c_{\mathrm{MAP}} := \langle m_0, T^*\Phi_{\mathrm{MAP}} - \varphi \rangle_{\mathcal{X}} \,.$$

Theorem [Kretschmann, Wachsmuth, Werner 2022]

Under a priori assumptions on $u^\dagger$, for every $\varphi \in \overline{\operatorname{ran} T^*}$, $\Phi \in \mathcal{Y}$, and $\alpha \in (0,1)$, rejection threshold $c = c(\varphi, \Phi, \alpha)$ can be chosen such that regularized test

$$\Psi_{\Phi,c}(Y) = \mathbf{1}_{\langle Y, \Phi \rangle > c}$$

has level α for testing H_0 against H_1.

Interpretation as regularized test

MAP test Ψ_{MAP} corresponds to regularized test $\Psi_{\Phi_{\mathrm{MAP}}, c_{\mathrm{MAP}}}$ with

$$c_{\mathrm{MAP}} := \langle m_0, T^*\Phi_{\mathrm{MAP}} - \varphi \rangle_{\mathcal{X}} .$$

Theorem [Kretschmann, Wachsmuth, Werner 2022]

Under a priori assumptions on $u^\dagger$, for every $\varphi \in \overline{\operatorname{ran} T^*}$, $\Phi \in \mathcal{Y}$, and $\alpha \in (0,1)$, rejection threshold $c = c(\varphi, \Phi, \alpha)$ can be chosen such that regularized test

$$\Psi_{\Phi,c}(Y) = \mathbf{1}_{\langle Y, \Phi \rangle > c}$$

has level α for testing H_0 against H_1.

MAP test Ψ_{MAP} has level α if prior mean m_0 is chosen according to

$$\langle m_0, T^*\Phi_{\mathrm{MAP}} - \varphi \rangle_{\mathcal{X}} = c(\varphi, \Phi_{\mathrm{MAP}}, \alpha).$$

Bonus slide: Optimality

Theorem [Kretschmann, Wachsmuth, Werner 2022]

For $\varphi \in \overline{\text{ran } T^*}$ and under a priori assumptions on $u^\dagger$, there exists optimal probe element $\Phi^\dagger \in \mathcal{Y}$ that maximizes power among all regularized level α tests.

Bonus slide: Optimality

Theorem [Kretschmann, Wachsmuth, Werner 2022]

For $\varphi \in \overline{\operatorname{ran} T^*}$ and under a priori assumptions on $u^\dagger$, there exists optimal probe element $\Phi^\dagger \in \mathcal{Y}$ that maximizes power among all regularized level α tests.

Theorem

If T is compact with singular system $(\tau_k, e_k, f_k)_{k\in\mathbb{N}}$ and if

$$\langle \varphi, e_k \rangle_{\mathcal{X}} = 0 \quad \text{for all } k \in \mathbb{N} \text{ with } \langle T^*\Phi^\dagger, e_k \rangle_{\mathcal{X}} = 0,$$

then prior covariance C_0 can be chosen such that power of Ψ_{MAP} is arbitrarily close to power of optimal regularized test $\Psi_{\Phi^\dagger, c(\varphi, \Phi^\dagger, \alpha)}$.

Structure

A priori assumptions on $u^\dagger$

Assumptions

1. Forward operator T is Hilbert–Schmidt and injective.
2. Spectral source condition

$$u^\dagger = (T^*T)^{\frac{\nu}{2}} w, \quad \|w\|_{\mathcal{X}} \leq \rho$$

   for some $w \in \mathcal{X}$ and $\nu, \rho > 0$.
3. Prior covariance operator

$$C_0 = \gamma^2 (T^*T)^\mu$$

   for some $\gamma > 0$ and $\mu \geq 1$.

A priori choice of prior covariance

Theorem (lower bound to power)

If $\mu > \frac{\nu}{2} - 1$, then power of Ψ_{MAP} is at least

$$\mathbb{P}_{u^\dagger}\left[\Psi_{\mathsf{MAP}}(Y) = 1\right] \geq Q\left(Q^{-1}(\alpha) + \frac{\frac{\langle \varphi, u^\dagger \rangle}{\|\varphi\|} - 2\rho\gamma^{-\frac{\nu}{\mu+1}}\sigma^{\frac{\nu}{\mu+1}}}{\gamma^{\frac{1}{\mu+1}}\sigma^{\frac{\mu}{\mu+1}}}\right).$$

A priori choice of prior covariance

Theorem (lower bound to power)

If $\mu > \frac{\nu}{2} - 1$, then power of Ψ_{MAP} is at least

$$\mathbb{P}_{u^\dagger}\left[\Psi_{\mathrm{MAP}}(Y) = 1\right] \geq Q\left(Q^{-1}(\alpha) + \frac{\frac{\langle\varphi, u^\dagger\rangle}{\|\varphi\|} - 2\rho\gamma^{-\frac{\nu}{\mu+1}}\sigma^{\frac{\nu}{\mu+1}}}{\gamma^{\frac{1}{\mu+1}}\sigma^{\frac{\mu}{\mu+1}}}\right).$$

Corollary (distinguishability)

If $\mu > \frac{\nu}{2} - 1$, then for any a priori choice

$$\gamma = \gamma_0 \sigma^{\omega}$$

with $\gamma_0 > 0$ and $\omega \in (-\mu, 1)$, power of Φ_{MAP} converges to 1 as $\sigma \to 0$.

- In the following, use constant a priori choice $\gamma = \gamma_0$.

Bonus slide: Separation rate

Corollary

Let $(u_\sigma^\dagger)_{\sigma>0}$ be a family in $\mathcal{X}$ that satisfies H_1,

$$\lim_{\sigma\to 0} \langle \varphi, u_\sigma^\dagger \rangle_{\mathcal{X}} = 0 \quad \text{and} \quad \lim_{\sigma\to 0} \frac{\langle \varphi, u_\sigma^\dagger \rangle_{\mathcal{X}}}{\sigma^{\frac{\nu}{\nu+1}}} = \infty.$$

If $\mu > \frac{\nu}{2} - 1$ and γ is chosen a priori as

$$\gamma = \sigma^{\frac{\nu-\mu}{\nu+1}},$$

then the power of Φ_{MAP} for $u_\sigma^\dagger$ converges to 1 as $\sigma \to 0$.

Oracle choice of prior covariance

- MAP test Ψ_{MAP} has power

$$\mathbb{P}_{u^\dagger}\left[\Psi_{\mathrm{MAP}}(Y) = 1\right] = Q\left(Q^{-1}(\alpha) - \frac{J_{Tu^\dagger}(\Phi_{\mathrm{MAP}})}{\sigma}\right),$$

with $J_{Tu^\dagger}\colon \mathcal{Y} \to \mathbb{R}$ [Kretschmann, Wachsmuth, Werner 2022].

Oracle MAP test

Choose $\gamma > 0$ in $C_0 = \gamma^2(T^*T)^\mu$ to maximize power of Ψ_{MAP}, i.e., as minimizer of

$$\gamma \mapsto J_{Tu^\dagger}(\Phi_{\mathrm{MAP}}(\gamma)).$$

- Objective functional $J_{Tu^\dagger}$ unaccessible.

A posteriori choice of prior covariance

- Use empirical objective functional J_Y instead of $J_{Tu^\dagger}$.

A posteriori MAP test

Choose $\gamma > 0$ in $C_0 = \gamma^2 (T^* T)^\mu$ as minimizer of

$$\gamma \mapsto J_Y(\Phi_{\mathrm{MAP}}(\gamma)) + \omega (\log \gamma)^2$$

for some $\omega > 0$.

- Due to dependence of Φ_{MAP} on Y via γ for this choice, it is no longer guaranteed that test has level α.

Numerical simulations

Considered problems

1. Deconvolution in 1D with kernel h,

$$(\mathcal{F}h)(\xi) = \left(1 + 0.06^2\xi^2\right)^{-2} \quad \text{for all } \xi \in \mathbb{R}.$$

2. Differentiation: Estimate second weak derivative of function $y \in H^2(0,1)$.
3. Backward heat equation on $(0,1)$ with Dirichlet boundary conditions.

- Choose truth $u^\dagger$ such that source condition is satisfied with $\nu = 1$.
- Choose prior smoothness $\mu = 1$.

Considered scenarios

- Construct a posteriori MAP test with nominal level $\tilde{\alpha} = 0.05$ and all other tests with level $\alpha = 0.1$.
- Compare power of different MAP tests with power of unregularized test in following two scenarios:

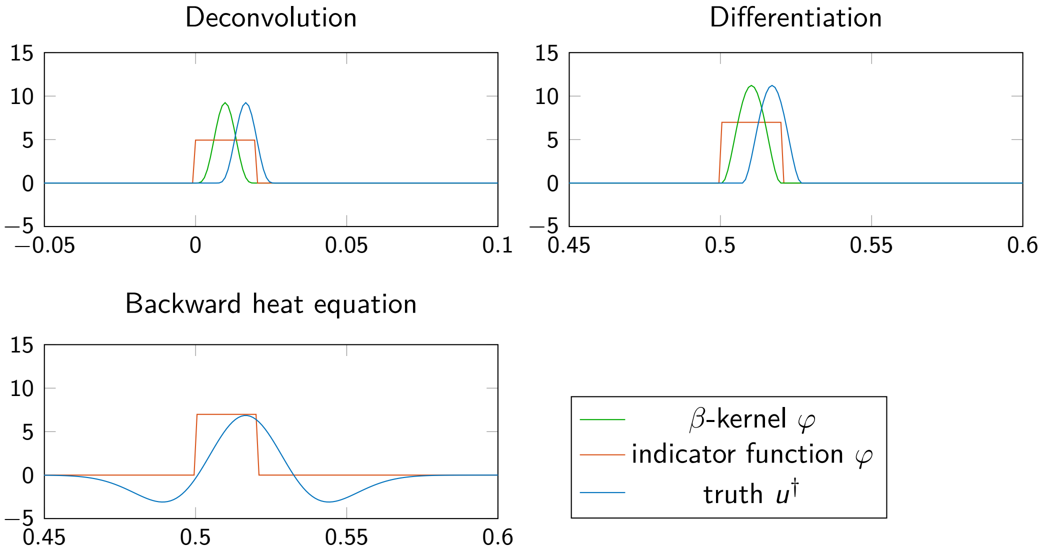
$\varphi \in \operatorname{ran} T^*$

- Choose φ as scaled β-kernel.
- Unregularized test well-defined.

$\varphi \notin \operatorname{ran} T^*$

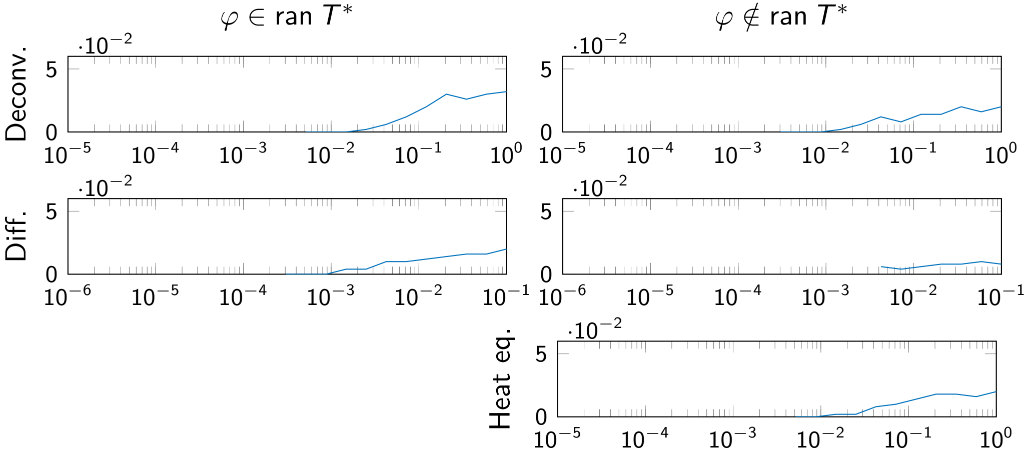
- Choose φ as indicator function.
- Unregularized test formally not defined.

Considered scenarios



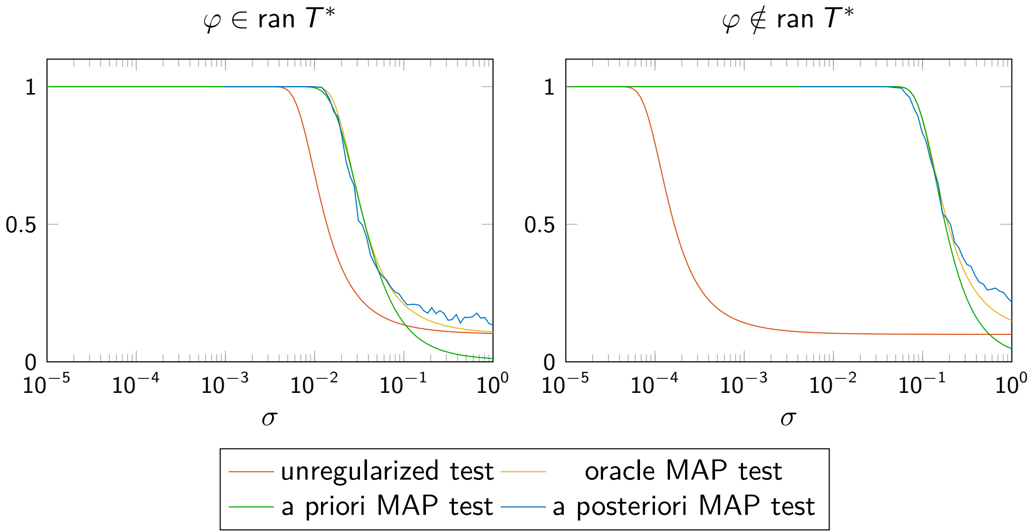
Numerical results

Empirical level of a posteriori MAP test remains below $\tilde{\alpha} = 0.05$ throughout all noise levels, problems, and scenarios.



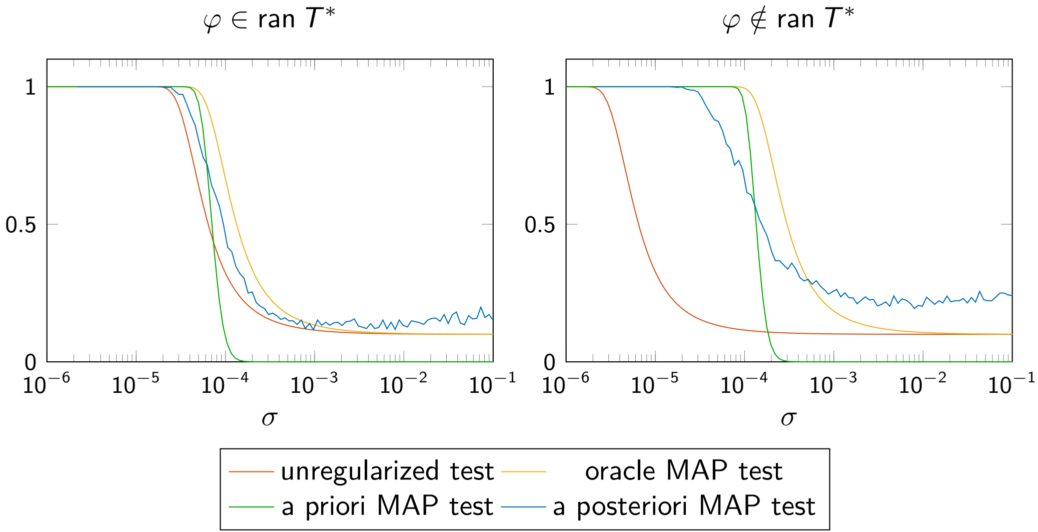
Numerical results: Deconvolution

Power of tests for different noise levels σ.



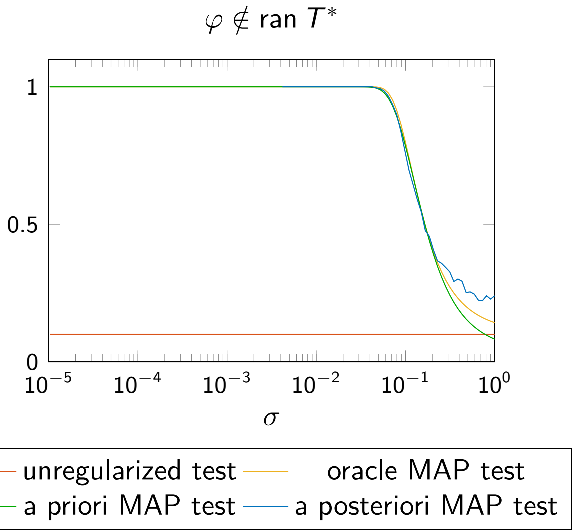
Numerical results: Differentiation

Power of tests for different noise levels σ.



Numerical results: Backward heat equation

Power of tests for different noise levels σ.



Conclusion

- MAP test based upon Gaussian prior can be evaluated via Tikhonov–Phillips regularization.
- MAP test is defined for any feature described by bounded linear functional $\varphi \in \mathcal{X}^*$.
- Regularizing effect allows feature testing in noise regimes where unregularized testing is unfeasible.

Outlook

- Construct MAP tests simultaneously for family of features.
- Other choices of prior distribution.
- Apply MAP tests to nonlinear inverse problems.

References

R. Kretschmann, F. Werner (2024).
Maximum a posteriori testing in statistical inverse problems.
Preprint, arXiv:2402.00686.

R. Kretschmann, D. Wachsmuth, F. Werner (2024).
Optimal regularized hypothesis testing in statistical inverse problems.
Inverse Problems 40, 015013, doi:10.1088/1361-6420/ad1132.

K. Proksch, F. Werner, A. Munk (2018).
Multiscale scanning in inverse problems.
Ann. Statist., 46(6B), doi:10.1214/17-AOS1669.