Asymptotic properties of multivariate tapering for estimation and prediction

Reinhard Furrer, François Bachoc, Juan Du

University of Zurich
University of Vienna
Kansas State University

Spatial Statistics: Emerging Patterns
June 2015, Avignon

Matrix manipulations in Kriging

- Consider a Gaussian process $Z(\boldsymbol{s})$ on $\mathbb{R}^d$:
  - zero mean
  - stationary covariance function $c(\boldsymbol{h}; \boldsymbol{\theta}_0) \in \{c(\boldsymbol{h}; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$
  - observed at $\boldsymbol{s}_1, ..., \boldsymbol{s}_n$

- Maximum Likelihood estimation : $\hat{\boldsymbol{\theta}}_{\mathrm{ML}} \in \mathrm{argmin}_{\boldsymbol{\theta}} L_{\boldsymbol{\theta}}$, with

$$L_{\boldsymbol{\theta}} = \frac{1}{n} \log\left(\det\left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right)\right) + \frac{1}{n} \boldsymbol{z}^{\mathsf{T}} \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \boldsymbol{z}.$$

  - observation vector $z = (Z(\boldsymbol{s}_1), ..., Z(\boldsymbol{s}_n))^{\mathsf{T}}$
  - $n \times n$ covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ with $\sigma_{\boldsymbol{\theta} ij} = c(\boldsymbol{s}_i - \boldsymbol{s}_j; \boldsymbol{\theta})$

- Kriging predictor of $Z(\boldsymbol{x})$:

$$\boldsymbol{\sigma}_{\boldsymbol{\theta}}(\boldsymbol{x})^{\mathsf{T}} \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \boldsymbol{z}$$

  - $\sigma_{\boldsymbol{\theta}}(\boldsymbol{x}) = (c(\boldsymbol{x} - \boldsymbol{s}_1; \theta), ..., c(\boldsymbol{x} - \boldsymbol{s}_n; \theta))^{\mathsf{T}}$

$\Longrightarrow$ Issues when n is large

Standard covariance functions are small but non-zero at large distance

- Ex. $c(\boldsymbol{h}; \theta) = e^{-(||\boldsymbol{h}||/\theta)}$

Covariance tapering : replacing the small covariances by zeros

$\Longrightarrow$ creates sparse covariance matrices $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ $\Longrightarrow$ faster linear algebra procedures. Ex. `R` package `SPAM`

The zeros are creating by replacing

$$c(\boldsymbol{h}; \boldsymbol{\theta}) \qquad \text{by} \qquad c(\boldsymbol{h}; \boldsymbol{\theta}) t(\boldsymbol{h}/\gamma)$$

- t : taper function, positive semi-definite with compact support and satisfying $t(\mathbf{0}) = 1$. Ex : Wendland1

$$t(\boldsymbol{h}) = (1 - \|\boldsymbol{h}\|)_+^4 (1 + 4\|\boldsymbol{h}\|)$$

- γ : taper range

- Tapered Maximum Likelihood estimation : $\hat{\boldsymbol{\theta}}_{tML} \in \mathrm{argmin}_{\boldsymbol{\theta}} \bar{L}_{\boldsymbol{\theta}}$, with

$$\bar{L}_{\boldsymbol{\theta}} = \frac{1}{n} \log\left(\det\left(\mathbf{K}_{\boldsymbol{\theta}}\right)\right) + \frac{1}{n} \boldsymbol{z}^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1} \boldsymbol{z}$$

  - $n \times n$ tapered covariance matrix $\mathbf{K}_{\boldsymbol{\theta}}$ with $k_{\boldsymbol{\theta} ij} = c(\boldsymbol{s}_i - \boldsymbol{s}_j; \boldsymbol{\theta}) t((\boldsymbol{s}_i - \boldsymbol{s}_j)/\gamma)$

- Tapered predictor of $Z(\boldsymbol{x})$:

$$\boldsymbol{k}_{\theta}(\boldsymbol{x})^\top \mathbf{K}_{\theta}^{-1} \boldsymbol{z}$$

  - $k_{\boldsymbol{\theta}}(\boldsymbol{x})_i = c(\boldsymbol{x} - \boldsymbol{s}_i; \theta) t((\boldsymbol{x} - \boldsymbol{s}_i)/\gamma)$

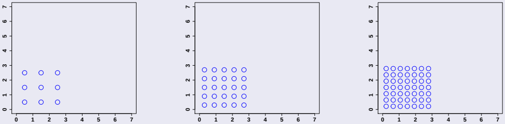
$\Longrightarrow$ Goal : studying the loss of accuracy caused by tapering

Asymptotic framework

Asymptotics : number of observations $n \to \infty$

Fixed-domain asymptotics

The observation points are dense in a bounded domain



In fixed-domain asymptotics, if

$$c(\boldsymbol{h};\boldsymbol{\theta}) \qquad \text{and} \qquad c(\boldsymbol{h};\boldsymbol{\theta})t(\boldsymbol{h}/\gamma)$$

have the same behavior at $\mathbf{0}$, then

$$\frac{\mathbb{E}\left[\left(\boldsymbol{\sigma}_{\boldsymbol{\theta}}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\boldsymbol{z} - Z(\boldsymbol{x})\right)^2\right]}{\mathbb{E}\left[\left(\boldsymbol{k}_{\boldsymbol{\theta}}(\boldsymbol{x})^{\mathsf{T}}\mathbf{K}_{\boldsymbol{\theta}}^{-1}\boldsymbol{z} - Z(\boldsymbol{x})\right)^2\right]} \to_{n\to\infty} 1$$

(Stein, 88, 90, 99, 13 ; Furrer et al. 06)

Results also exist for Maximum Likelihood : Du et al. 09

Multivariate case

- p inter-correlated Gaussian processes on $\mathbb{R}^d$

$$Z_1(\boldsymbol{s}), ..., Z_p(\boldsymbol{s})$$

  - All observed at $\boldsymbol{s}_1, ..., \boldsymbol{s}_n$

- Covariance and cross-covariance functions

$$\{c_{kl}(\boldsymbol{h}, \boldsymbol{\theta})\} \qquad \text{with} \qquad c_{kl}(\boldsymbol{h}, \boldsymbol{\theta}_0) = \mathrm{Cov}\left[Z_k(\boldsymbol{s}+\boldsymbol{h}), Z_l(\boldsymbol{s})\right]$$

- Taper functions

$$\{t_{kl}(\boldsymbol{h}/\gamma)\}$$

- New matrix and vector quantities :
  - Observation vector $\boldsymbol{z}$ of size $np \times 1$ filled as

$$(Z_1(\boldsymbol{s}_1), ..., Z_1(\boldsymbol{s}_n), ..., Z_p(\boldsymbol{s}_1), ..., Z_p(\boldsymbol{s}_n))^T$$

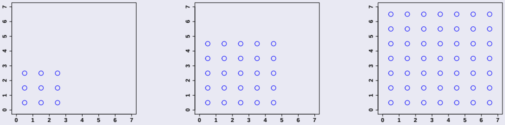
  - Prediction of $Z_1(\boldsymbol{x})$
  - $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, $\mathbf{K}_{\boldsymbol{\theta}}$, $\boldsymbol{\sigma}_{\boldsymbol{\theta}}(\boldsymbol{x})$, $\boldsymbol{k}_{\boldsymbol{\theta}}(\boldsymbol{x})$ filled appropriately

$\Longrightarrow$ Same equations for Maximum Likelihood and prediction

Theoretical tools of the univariate case for fixed-domain asymptotics are (to our knowledge) not available in the multivariate case

Increasing-domain asymptotics

A minimum spacing exists between the observation points $\longrightarrow$ infinite observation domain.



Increasing-domain asymptotics also studied by Shaby and Ruppert, 2012 (univariate), Bevilacqua et al., 2015 (multivariate)

Additional benefit : more general results

Our assumptions

- $c_{kl}(\boldsymbol{x};\boldsymbol{\theta})$ is continuously differentiable with respect to $\boldsymbol{\theta}$. There exist $A < +\infty$ and $\alpha > 0$ so that,

$$|c_{kl}(\boldsymbol{x};\boldsymbol{\theta})| \leq \frac{A}{1+|\boldsymbol{x}|^{d+\alpha}} \quad \text{and} \quad \left|\frac{\partial}{\partial\theta_i} c_{kl}(\boldsymbol{x};\boldsymbol{\theta})\right| \leq \frac{A}{1+|\boldsymbol{x}|^{d+\alpha}}$$

  - Standard for infinitely-supported covariance functions

- The taper functions t_{kl} are continuous at $\mathbf{0}$ and satisfy $t_{kl}(\mathbf{0}) = 1$ and $|t_{kl}(\boldsymbol{x})| \leq 1$. The taper range $\gamma = \gamma_n$ satisfies $\gamma_n \to_{n\to\infty} +\infty$
  - No rate assumed

- There exists $\Delta > 0$ so that for all $i \neq j$, $||\boldsymbol{s}_i - \boldsymbol{s}_j|| \geq \Delta$
  - (increasing-domain asymptotics)

We show :

Theorem

As $n \to \infty$,

$$\sup_{\theta \in \Theta} |L_\theta - \bar{L}_\theta| = o_p(1).$$

Corollary

If Maximum Likelihood is consistent, then tapered Maximum Likelihood is consistent

Comments

- Without rate assumptions on taper range γ we preserve consistency
- We would need rate assumptions to preserve a $\sqrt{n}$ rate of convergence
- Analysis is different for the two-taper Maximum Likelihood : Shaby and Ruppert 2012 (univariate), Bevilacqua et al., 2015 (multivariate)

Theorem

Let $\mathcal{D}_n$ be a sequence of measurable subsets of $\mathbb{R}^d$ with positive Lebesgue measures and let $f_n(\boldsymbol{x})$ be a sequence of continuous probability density functions on $\mathcal{D}_n$. Then, as $n \to \infty$,

$$\sup_{\boldsymbol{\theta} \in \Theta} \left| \int_{\mathcal{D}_n} \left[\boldsymbol{\sigma}_{\boldsymbol{\theta}}(\boldsymbol{x})^{\mathsf{T}} \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \boldsymbol{z} - Z_1(\boldsymbol{x}) \right]^2 f_n(\boldsymbol{x}) d\boldsymbol{x} - \int_{\mathcal{D}_n} \left[\boldsymbol{k}_{\boldsymbol{\theta}}(\boldsymbol{x})^{\mathsf{T}} \mathbf{K}_{\boldsymbol{\theta}}^{-1} \boldsymbol{z} - Z_1(\boldsymbol{x}) \right]^2 f_n(\boldsymbol{x}) d\boldsymbol{x} \right|$$

goes to 0 in probability

- Errors are lower-bounded $\Longrightarrow$ relative efficiency
- Also when predicting with $\hat{\theta}$

Matérn direct and cross-covariance functions :

$$c_{kl}(\boldsymbol{x};\boldsymbol{\theta}) = \frac{\sigma_{kl}^2}{2^{\nu_{kl}-1}\Gamma(\nu_{kl})}(\|\boldsymbol{x}\|/\rho_{kl})^{\nu_{kl}}\mathcal{K}_{\nu_{kl}}(\|\boldsymbol{x}\|/\rho_{kl}), \qquad k,l=1,2$$

Estimation of $\boldsymbol{\theta} = (\rho_{11},\rho_{12},\rho_{22},\sigma_{11},\sigma_{12},\sigma_{22})^{\mathrm{T}}$:

1. ranges : $\rho_{11}=5$, $\rho_{12}=3$, $\rho_{22}=4$
   sills : $\sigma_{11}=1$, $\sigma_{12}=.6$, $\sigma_{22}=1$
   smoothness : $\nu_{11}=\nu_{12}=\nu_{22}=1/2$
2. ranges : $\rho_{11}=3$, $\rho_{12}=3$, $\rho_{22}=4$
   sills : $\sigma_{11}=1$, $\sigma_{12}=.7$, $\sigma_{22}=1$
   smoothness : $\nu_{11}=3/2$, $\nu_{12}=1$, $\nu_{22}=1/2$

Taper matrix function :

1. $t_{kl}(\boldsymbol{x}) = (1-\|\boldsymbol{x}\|)_+^4(1+4\|\boldsymbol{x}\|), \qquad k,l=1,2.$
2. $t_{kl}(\boldsymbol{x}) = (1-\|\boldsymbol{x}\|)_+^6(1+6\|\boldsymbol{x}\|+35\|\boldsymbol{x}\|^2/3), \qquad k,l=1,2.$
3. $t_{kl}(\boldsymbol{x}) = (1-\|\boldsymbol{x}\|)_+^2(1+\|\boldsymbol{x}\|/2), \qquad k,l=1,2.$
4. $t_{11}(\boldsymbol{x}) = (1-\|\boldsymbol{x}\|)_+^5(1+5\|\boldsymbol{x}\|+\|\boldsymbol{x}\|^2),$
   $t_{12}(\boldsymbol{x}) = t_{21}(\boldsymbol{x}) = \sqrt{6/7}\,(1-\|\boldsymbol{x}\|)_+^5(1+5\|\boldsymbol{x}\|+\|\boldsymbol{x}\|^2),$
   $t_{22}(\boldsymbol{x}) = (1-\|\boldsymbol{x}\|)_+^5(1+5\|\boldsymbol{x}\|).$

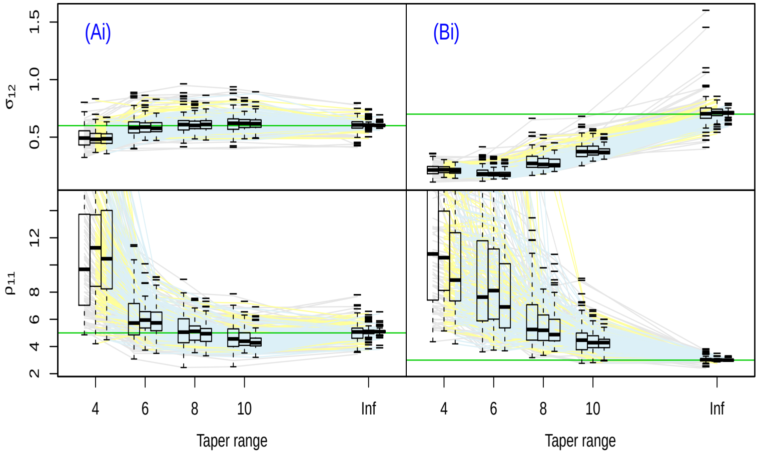
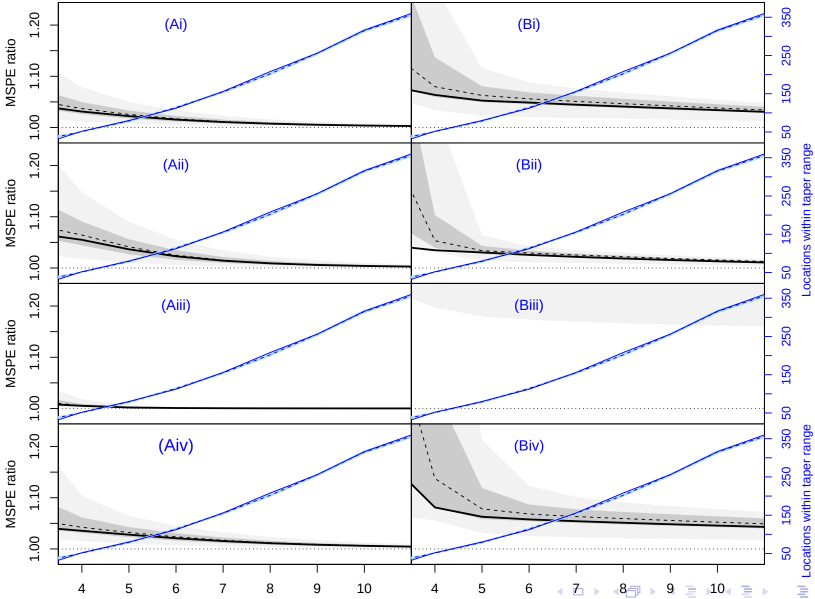


FIGURE: Effect of increasing the domain on the ML estimates. The boxplots correspond to $n = 400$ (gray), 1024 (yellow), 2500 (light blue), left to right for each taper range.

Ratios of prediction mean square errors for $n = 400$



Conclusion

Conclusion :

- Kriging equations require handling $n \times n$ matrices
- Covariance tapering is one convenient approximation (among others)
- We have addressed the statistical error in the multivariate case
- Tapering appears particularly attractive for prediction

Some open questions :

- Rate on taper range for rate of Maximum Likelihood
- Fixed-domain asymptotic analysis

The paper :

R. Furrer, F. Bachoc, J. Du (2015+). Asymptotic properties of multivariate tapering for estimation and prediction, **`http://arxiv.org/abs/1506.01833`**

Thank you for your attention !