X x X X

MMMMMMM YYYY, Volume VV, Issue II. Author

Facile Spacio-Temporal Estimation

Fixed Rank Kriging with Adaptive Least Squares

Dave Zes
Department of Statistics, UCLA

Abstract

In the following we propose a simple temporal extension to fixed rank kriging. We call upon variants of adaptive least squares to provide the necessary time-adapting estimators/predictors. Simulations provided.

Keywords: fixed rank kriging, kriging, adaptive least squares, covariance function, kalman filter.

1. Introduction

Fixed rank kriging (FRK) represents yet another excellent contribution to the science of spacio-temporal modeling by Noel Cressie along with Gardar Johannesson, Cressie, Noel and Johannesson, Gardar (2008). FRK is a response to the growing need to analyze or predict scalar-valued observations in space where their number, n, is very large, say $n > 1000$.

FRK bypasses the usual requirement of the inversion of the $n \times n$ predictor covariance matrix, $\mathbf{\Sigma}$. Cressie and Johannesson obviate this burden by developing three key insights. First, they have $\mathbf{\Sigma} \approx \mathbf{SKS}^T$, with $\mathbf{S}$ an $n \times r$ (with $r << n$) matrix containing distance information between the n sites and r fixed *centers*, and $\mathbf{K}$ an $r \times r$ matrix estimated empirically. Second, they bin up the observations, $\mathbf{z}$ into $\bar{\mathbf{z}}$, with length m, with $r < m < n$. And lastly they invoke the Sherman-Morrison-Woodbury equality.

In this current submission we offer no further detailing of FRK; we shall assume the reader possesses facility with this important method. A thorough understanding of ALS, on the other hand, is not necessarily required, since we make its implementation explicit, e.g., (1), (2), (6), (9), (10).

2. Fixed Rank Kriging with Adaptive Least Squares

Adaptive least squares provides a means of estimating a potentially time-varying linear function between a predictor and response through recursive adjusting of the predictor-predictor and predictor-response covariance matrices, Haykin (2002), Sayed (2003), McCulloch (2005). FRK allows for the natural insertion of ALS in two places. First, for the sake of estimating/predicting the mean function $\widehat{\boldsymbol{\alpha}}_t$, and second, estimating/predicting the binned de-trended response variance matrix, $\bar{\boldsymbol{\Sigma}}_{M\,t}$. Our proposed ALS extension is made yet more desirable because many FRK computations can be executed prior to the process loop, e.g., the creation of $\mathbf{S}$, $\mathbf{S}_0$, $\mathbf{W}$, as well as the QR decomposition of $\mathbf{S}$, to name a few.

2.1. 0-Step-Ahead Prediction

Here we suppose we have been charged with producing $\widehat{\mathbf{y}}_{0\,t}|\mathbf{Z}_{1:t}$. The following 11 assignments are recursed over time.

$$\widehat{\mathbf{L}}_{\mathrm{TT}\,t} = \widehat{\mathbf{L}}_{\mathrm{TT}\,t-1} + g_{1\,t} \cdot \left(\mathbf{T}_t^T\, \widehat{\boldsymbol{\Sigma}}_{t-1}^{-1}\, \mathbf{T}_t - \widehat{\mathbf{L}}_{\mathrm{TT}\,t-1}\right) \tag{1}$$

$$\widehat{\mathbf{L}}_{\mathrm{Tz}\,t} = \widehat{\mathbf{L}}_{\mathrm{Tz}\,t-1} + g_{1\,t} \cdot \left(\mathbf{T}_t^T\, \widehat{\boldsymbol{\Sigma}}_{t-1}^{-1}\, \mathbf{z}_t - \widehat{\mathbf{L}}_{\mathrm{Tz}\,t-1}\right) \tag{2}$$

$$\widehat{\boldsymbol{\alpha}}_t = \left(\widehat{\mathbf{L}}_{\mathrm{TT}\,t} + \lambda \mathbf{I}\right)^{-1} \widehat{\mathbf{L}}_{\mathrm{Tz}\,t} \tag{3}$$

$$\widetilde{\mathbf{z}}_t = \mathbf{z}_t - \mathbf{T}_t\, \widehat{\boldsymbol{\alpha}}_t \tag{4}$$

$$\bar{\mathbf{z}}_t = \mathbf{W}^T\, \widetilde{\mathbf{z}}_t\, * \boldsymbol{\gamma} \tag{5}$$

$$\bar{\boldsymbol{\Sigma}}_{M\,t} = \bar{\boldsymbol{\Sigma}}_{M\,t-1} + g_{2\,t} \cdot \left(\mathsf{Var}[\bar{\mathbf{z}}_t] - \bar{\boldsymbol{\Sigma}}_{M\,t-1}\right) \tag{6}$$

$$\widehat{\mathbf{K}}_t = \mathbf{R}^{-1}\mathbf{Q}^T \left(\bar{\boldsymbol{\Sigma}}_{M\,t} - \sigma^2 \bar{\mathbf{V}}\right) \mathbf{Q} \left(\mathbf{R}^{-1}\right)^T \tag{7}$$

$$\widehat{\boldsymbol{\Sigma}}_t^{-1} = (\sigma^2\mathbf{V})^{-1} - (\sigma^2\mathbf{V})^{-1}\mathbf{S} \left(\widehat{\mathbf{K}}_t^{-1} + \mathbf{S}^T\, (\sigma^2\mathbf{V})^{-1}\mathbf{S}\right)^{-1} \mathbf{S}^T\, (\sigma^2\mathbf{V})^{-1} \tag{8}$$

$$g_{1\,t+1} = (g_{1\,t} + \rho_1)/(g_{1\,t} + \rho_1 + 1) \tag{9}$$

$$g_{2\,t+1} = (g_{2\,t} + \rho_2)/(g_{2\,t} + \rho_2 + 1) \tag{10}$$

$$\widehat{\mathbf{y}}_{0\,t} = \mathbf{T}_0\, \widehat{\boldsymbol{\alpha}}_t + \mathbf{S}_0\, \widehat{\mathbf{K}}_t\, \mathbf{S}^T\, \widehat{\boldsymbol{\Sigma}}_t^{-1}\, \widetilde{\mathbf{z}}_t \tag{11}$$

Our process requires 3 scalar hyperparameters, $(\lambda, \rho_1, \rho_2)$. Equations (1), (2) are ALS updates of the predictor-predictor variance and the predictor-response covariance matrices respectively; (4) de-trends the response; (5) creates the binned response, with $\boldsymbol{\gamma}$ the reciprocal of the column sums of $\mathbf{W}$; (6) creates an ALS update of the binned response variance matrix. Finally, (9) and (10) update the update gains.

In absence of any prior conviction concerning the initial state of the three covariance matrices, a sensible way to commence the algorithm is to have $g_{1\,1} = g_{2\,1} = 1$, and $\widehat{\mathbf{L}}_{\mathrm{TT}\,0} = \mathbf{0}$, $\widehat{\mathbf{L}}_{\mathrm{Tz}\,0} = \mathbf{0}$, $\bar{\boldsymbol{\Sigma}}_{M\,0} = \mathbf{0}$, $\widehat{\boldsymbol{\Sigma}}_0^{-1} = \mathbf{I}$.

2.2. 1-Step-Ahead Prediction

Here we create $\widehat{\mathbf{y}}_{0\,t} \mid \mathbf{Z}_{1:(t-1)}$.

$$\widehat{\mathbf{L}}_{\mathrm{TT}\,t} = \widehat{\mathbf{L}}_{\mathrm{TT}\,t-1} + g_{1\,t} \cdot \left(\mathbf{T}_t^T \, \widehat{\boldsymbol{\Sigma}}_{t-1}^{-1} \, \mathbf{T}_t - \widehat{\mathbf{L}}_{\mathrm{TT}\,t-1}\right) \tag{12}$$

$$\widehat{\mathbf{L}}_{\mathrm{Tz}\,t} = \widehat{\mathbf{L}}_{\mathrm{Tz}\,t-1} + g_{1\,t} \cdot \left(\mathbf{T}_t^T \, \widehat{\boldsymbol{\Sigma}}_{t-1}^{-1} \, \mathbf{z}_t - \widehat{\mathbf{L}}_{\mathrm{Tz}\,t-1}\right) \tag{13}$$

$$\widehat{\boldsymbol{\alpha}}_t = \left(\widehat{\mathbf{L}}_{\mathrm{TT}\,t} + \lambda \mathbf{I}\right)^{-1} \widehat{\mathbf{L}}_{\mathrm{Tz}\,t} \tag{14}$$

$$\widetilde{\mathbf{z}}_t = \mathbf{z}_t - \mathbf{T}_t \, \widehat{\boldsymbol{\alpha}}_t \tag{15}$$

$$\bar{\mathbf{z}}_t = \mathbf{W}^T \, \widetilde{\mathbf{z}}_t \, \ast \boldsymbol{\gamma} \tag{16}$$

$$\bar{\boldsymbol{\Sigma}}^{(0)}_{M\,t} = \bar{\boldsymbol{\Sigma}}^{(0)}_{M\,t-1} + g_{2\,t} \cdot \left(\mathsf{Var}[\bar{\mathbf{z}}_t] - \bar{\boldsymbol{\Sigma}}^{(0)}_{M\,t-1}\right) \tag{17}$$

$$\bar{\boldsymbol{\Sigma}}^{(1)}_{M\,t} = \bar{\boldsymbol{\Sigma}}^{(1)}_{M\,t-1} + g_{2\,t} \cdot \left(\mathsf{Cov}[\bar{\mathbf{z}}_{t-1}, \bar{\mathbf{z}}_t] - \bar{\boldsymbol{\Sigma}}^{(1)}_{M\,t-1}\right) \tag{18}$$

$$\widehat{\mathbf{K}}^{(0)}_t = \mathbf{R}^{-1}\mathbf{Q}^T \left(\bar{\boldsymbol{\Sigma}}^{(0)}_{M\,t} - \sigma^2 \bar{\mathbf{V}}\right) \mathbf{Q} \left(\mathbf{R}^{-1}\right)^T \tag{19}$$

$$\widehat{\mathbf{K}}^{(1)}_t = \mathbf{R}^{-1}\mathbf{Q}^T \, \bar{\boldsymbol{\Sigma}}^{(1)}_{M\,t} \, \mathbf{Q} \left(\mathbf{R}^{-1}\right)^T \tag{20}$$

$$\widehat{\boldsymbol{\Sigma}}_t^{-1} = \left(\sigma^2 \mathbf{V}\right)^{-1} - \left(\sigma^2 \mathbf{V}\right)^{-1} \mathbf{S} \, \left(\left(\widehat{\mathbf{K}}^{(0)}_t\right)^{-1} + \mathbf{S}^T \, (\sigma^2 \mathbf{V})^{-1} \mathbf{S}\right)^{-1} \mathbf{S}^T \, \left(\sigma^2 \mathbf{V}\right)^{-1} \tag{21}$$

$$g_{1\,t+1} = (g_{1\,t} + \rho_1)/(g_{1\,t} + \rho_1 + 1) \tag{22}$$

$$g_{2\,t+1} = (g_{2\,t} + \rho_2)/(g_{2\,t} + \rho_2 + 1) \tag{23}$$

$$\widehat{\mathbf{y}}_{0\,t} = \mathbf{T}_{0\,t-1} \, \widehat{\boldsymbol{\alpha}}_{t-1} + \mathbf{S}_0 \, \widehat{\mathbf{K}}^{(1)}_{t-1} \, \mathbf{S}^T \, \widehat{\boldsymbol{\Sigma}}_{t-2}^{-1} \, \widetilde{\mathbf{z}}_{t-1} \tag{24}$$

3. Simulations

For the sake of illustration we created data using the state-space system:

$$\boldsymbol{\alpha}_t = \mathbf{F} \, \boldsymbol{\alpha}_{t-1} + \boldsymbol{\nu}_t \tag{25}$$

$$\boldsymbol{z}_t^T = \mathbf{T} \, \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t \tag{26}$$

with $\boldsymbol{\nu}_t \sim \mathcal{N}[\,\mathbf{0}, \boldsymbol{\Upsilon}\,]$ and $\boldsymbol{\varepsilon}_t \sim \mathcal{N}[\,\mathbf{0}, \boldsymbol{\Sigma}\,]$, and $t \in 1\!:\!\tau$.

We have our spacial support be the *ring* (modulo 10) $\Omega = [10, 20] \times [0, 10] \subset \mathbb{R}^2$, the ith site notated with the tuple $s_i = (x_i, y_i)$, and have our mean function be a tensor product cosine expansion over Ω.

$$\mathbf{T}_{i\cdot} = \mathbf{T}_{i\cdot}(\boldsymbol{\omega}, \boldsymbol{\upsilon}, s_i) = \left(\cos\left[\omega_1 x_i\right], ..., \cos\left[\omega_{k_x} x_i\right]\right) \otimes \left(\cos\left[\upsilon_1 y_i\right], ..., \cos\left[\upsilon_{k_y} y_i\right]\right) \tag{27}$$

with $\otimes$ operating on vectors of length k_x and k_y and returning a vector of length $k_x \cdot k_y$; $\boldsymbol{\omega}$, $\boldsymbol{\upsilon}$ are vectors of frequencies on the longitude and latitude respectively of length k_x and k_y.

In each example, we estimated the three process hyperparameters using Metropolis-Hastings minimizing cross-validation RMSE for 24 randomly chosen sites over $t \in 100\!:\!\tau$

In the movies, observations are shown as circles, diameter proportional to magnitude. Blue dots show bin centers, green dots show multi-resolutional centers. A raster heat map shows spacial prediction (red, low value; white, high value).

3.1. Example 1, 196 Sites

0-Step-Ahead Prediction:

n	196
τ	1000
$\boldsymbol{\omega}$	$(0.2\pi, 0.4\pi)$
$\boldsymbol{\upsilon}$	$(0.2\pi, 0.4\pi)$
Σ	$\exp\{-0.4 \cdot \mathbf{D}\} + 0.5\mathbf{I}$
Υ	$0.1 \cdot \mathbf{I}$
$\mathbf{F}$	$0.9999 \cdot \mathbf{I}$
$\boldsymbol{\alpha}_0$	$(7, -1, 7, -1)^T$

Table 1: Simulation Parameters

r	21
m	49
$\mathsf{range}[\mathbf{Z}]$	$(-31.71, 29.37)$
ρ_1	5.081
ρ_2	0.03170
λ	$3.361 \cdot 10^{-5}$
RMSE	1.0874

Table 2: Stats, Hyperparameters, Results

❇ **Movie, $n = 196$, 0-Step-Ahead** ❇

1-Step-Ahead Prediction:

ρ_1	1.484
ρ_2	$9.797 \cdot 10^{-6}$
λ	$8.677 \cdot 10^{-5}$
RMSE	1.277

Table 3: Stats, Hyperparameters, Results

❇ **Movie, $n = 196$, 1-Step-Ahead** ❇

3.2. Example 2, 784 Sites

0-Step-Ahead Prediction:

n	784
τ	1000
$\boldsymbol{\omega}$	$(0.2\pi, 0.4\pi, 0.6\pi)$
$\boldsymbol{\upsilon}$	$(0.2\pi, 0.4\pi)$
Σ	$\exp\{-0.4 \cdot \mathbf{D}\} + 0.5\mathbf{I}$
Υ	$0.1 \cdot \mathbf{I}$
$\mathbf{F}$	$0.9999 \cdot \mathbf{I}$
$\boldsymbol{\alpha}_0$	$(7, -1, 7, -1, 0, 0)^T$

Table 4: Simulation Parameters

r	21
m	324
$\mathsf{range}[\mathbf{Z}]$	$(-38.03, 31.79)$
ρ_1	36.84
ρ_2	0.2888
λ	$8.726 \cdot 10^{-6}$
RMSE	1.0270

Table 5: Stats, Hyperparameters, Results

❇ **Movie, $n = 784$, 0-Step-Ahead** ❇

1-Step-Ahead Prediction:

ρ_1	2.013
ρ_2	$7.054 \cdot 10^{-6}$
λ	$8.758 \cdot 10^{-5}$
RMSE	1.299

Table 6: Stats, Hyperparameters, Results

❊ **Movie, $n = 784$, 1-Step-Ahead** ❊

4. Final Thoughts

The union between FRK and ALS appears quite natural, making massive space-time data tractable. For the sake of brevity we examined only 2 simulated data, but preliminary experimentation shows FRKALS to be robust and generally indifferent to selection of mean function and system parameterization.

References

Cressie, Noel, Johannesson, Gardar (2008). “Fixed rank kriging for very large spatial data sets.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, pp. 209–226.

Haykin S (2002). *Adaptive Filter Theory.* Forth edition. Prentice-Hall, Upper Saddle River.

McCulloch JH (2005). “The Kalman Foundations of Adaptive Least Squares, With Application to U.S. Inflation.” *Unpublished.*

Sayed AH (2003). *Fundamentals of Adaptive Fileterimg.* John Wiley & Sons, Hoboken, N.J.

Affiliation:
Dave Zes
Department of Statistics, UCLA
Los Angeles, CA 90095-1554
E-mail: `davezes@stat.ucla.edu`
URL: `http://www.stat.ucla.edu/~davezes`

XXX

Volume VV, Issue II
MMMMMM YYYY

Submitted: yyyy-mm-dd
Accepted: yyyy-mm-dd