Supplementary Materials for *GLS Kernel Regression for Network-Structured Data*

Edward Antonian, Gareth W. Peters, Mike Chantler, and Hongxuan Yan

I. Table of identities

The following is a table of metamathematical identities used in various places within this paper to derive expressions. The equations listed in this section are considered well-known or easily verifiable, and are therefore provided without proof. The majority can be found in standard references such as [1] and [2].

TABLE I
A series of identities involving Matrices, Kronecker products and vectorization.

	Expression	Alternative expression
1	$\text{vec}(AXB)$	$(B^\top \otimes A)\,\text{vec}(X)$
2	$\text{tr}\big(A^\top B\big)$	$\text{vec}(A)^\top \text{vec}(B)$
3	$AC \otimes BD$	$(A \otimes B)(C \otimes D)$
4	$(A \otimes B)^{-1}$	$A^{-1} \otimes B^{-1}$
5	$\text{tr}\big(X^\top AYB\big)$	$\text{vec}(X)^\top (B^\top \otimes A)\,\text{vec}(Y)$
6	$\text{vec}(J \circ Y)$	$\text{diag}\big(\text{vec}(J)\big)\text{vec}(Y)$
7	$\ln(\det(A \otimes B))$	$\text{rank}(B)\ln\lvert A\rvert + \text{rank}(A)\ln\lvert B\rvert$
8	$(B^\top B + A^{-1})^{-1}B^\top$	$A^{-1}B^\top(I + BAB^\top)^{-1}$
9	$\text{diag}^{-1}\big(A\,\text{diag}(\mathbf{x})\,B\big)$	$(B^\top \circ A)\,\mathbf{x}$

TABLE II
The matrix derivative of several useful scalar functions.

	$f(X)$	$\partial f / \partial X$
1	$\text{tr}\big((A - BX)^\top (A - BX)\big)$	$-2B^\top A + 2B^\top BX$
2	$\text{tr}\big(XAX^\top\big)$	$X(A + A^\top)$
3	$\text{tr}\big(AXB\big)$	$A^\top B^\top$
4	$\text{tr}\big(AXBX^\top C\big)$	$A^\top C^\top XB^\top + CAXB$
5	$\text{tr}\big(AX^{-1}B\big)$	$-(X^{-1}BAX^{-1})^\top$
6	$\ln\lvert\det(X)\rvert$	$X^{-\top}$

II. Proofs and derivations

Theorem II.1. *For a cost function*

$$\xi(F) = \text{tr}\big((Y - S_N F S_T^\top)^\top (Y - S_N F S_T^\top)\big) + \gamma\,\text{tr}\big(K^{-1}F^\top H^{-2} F\big)$$

the minimising value of F is given by

$$F^\star = H^2 S_N^\top \bar{U}\big(J \circ (\bar{U}^\top Y \bar{V})\big)\bar{V}^\top S_T K$$

where

$$S_T^\top K S_T = \bar{V}\bar{\Lambda}_K\bar{V}^\top, \quad \textit{and} \quad S_N^\top H^2 S_N = \bar{U}\bar{\Lambda}_H\bar{U}^\top$$

and $J \in \mathbb{R}^{N' \times T'}$ has the elements given by

$$J_{ij} = \frac{1}{\bar{\lambda}_j^K \bar{\lambda}_i^H + \gamma}$$

Proof. The derivative of ξ with respect to F is given by

$$\frac{\partial \xi}{\partial F} = 2S_N^\top(S_N F S_T^\top - Y)S_T^\top + 2\gamma H^{-2} F K^{-1}$$

Setting this equal to zero an rearranging gives an equation for $F^\star$.

$$S_N^\top S_N F^\star S_T^\top S_T + \gamma H^{-2} F^\star K^{-1} = S_N Y S_T^\top$$

Vectorizing both sides and applying identity 1 from table 1 gives:

$$\text{vec}(F^\star) = \Big(S_T^\top S_T \otimes S_N^\top S_N + \gamma K^{-1} \otimes H^{-2}\Big)^{-1} \big(S_T^\top \otimes S_N^\top\big)\text{vec}(Y)$$

Applying identity 8 from table 1 means this can be rewritten as follows:

$$= \big(K S_T \otimes H^2 S_N\big) \times \Big(\gamma I_{T'} \otimes I_{N'} + S_T^\top K S_T \otimes S_N^\top H^2 S_N\Big)^{-1}\text{vec}(Y)$$

Next, substitute in the eigenvalue decompositions.

$$= \big(K S_T^\top \otimes H^2 S_N^\top\big) \times \Big(\gamma I_{T'} \otimes I_{N'} + \bar{V}\bar{\Lambda}_K\bar{V}^\top \otimes \bar{U}\bar{\Lambda}_H\bar{U}^\top\Big)^{-1}\text{vec}(Y)$$

$$= \big(K S_T^\top \otimes H^2 S_N^\top\big) \times \Big(\big(\bar{V} \otimes \bar{U}\big)\big(\gamma I_{T'} \otimes I_{N'} + \bar{\Lambda}_K \otimes \bar{\Lambda}_H\big)\big(\bar{V}^\top \otimes \bar{U}^\top\big)\Big)^{-1}\text{vec}(Y)$$

$$= \big(K S_T^\top \otimes H^2 S_N^\top\big)\big(\bar{V} \otimes \bar{U}\big) \times \big(\gamma I_{T'} \otimes I_{N'} + \bar{\Lambda}_K \otimes \bar{\Lambda}_H\big)\big(\bar{V}^\top \otimes \bar{U}^\top\big)\text{vec}(Y)$$

Note that the matrix $\big(\gamma I_{T'} \otimes I_{N'} + \bar{\Lambda}_K \otimes \bar{\Lambda}_H\big)$ is diagonal. This means it can be rewritten as $\operatorname{diag}\big(\operatorname{vec}(J)\big)$ where J is defined above.

$$= \big(K S_T^\top \bar{V} \otimes H^2 S_N^\top \bar{U}\big) \operatorname{diag}\big(\operatorname{vec}(J)\big) \operatorname{vec}(\bar{U}^\top Y \bar{V})$$

Finally, applying identities 1 and 6 gives the desired expression.

□

Theorem II.2. *For a cost function*

$$\xi(F) = \operatorname{tr}\big((Y - S_N F S_T^\top)^\top \Sigma_N^{-1} (Y - S_N F S_T^\top)\, \Sigma_T^{-1}\big) \\ + \gamma \operatorname{tr}\big(K^{-1} F^\top H^{-2} F\big)$$

The minimising value for Ψ can be expressed as

$$F^\star = B\,(J \circ \bar{Y})\, C^\top$$

where

$$\Sigma_T = \Psi \Lambda_{\Sigma_T} \Psi^\top, \quad \Sigma_N = \Phi \Lambda_{\Sigma_N} \Phi^\top$$

$$\Lambda_{\Sigma_T}^{-1/2} \Psi^\top S_T K S_T^\top \Psi \Lambda_{\Sigma_T}^{-1/2} = \bar{V} \bar{\Lambda}_K \bar{V}^\top \\ \Lambda_{\Sigma_N}^{-1/2} \Phi^\top S_N H^2 S_N^\top \Phi \Lambda_{\Sigma_N}^{-1/2} = \bar{U} \bar{\Lambda}_H \bar{U}^\top$$

and

$$J_{ij} = \frac{1}{\gamma + \bar{\lambda}_j^K \bar{\lambda}_i^H} \\ B = H^2 S_N^\top \Phi \Lambda_{\Sigma_N}^{-1/2} \bar{U} \\ C = K S_T^\top \Psi \Lambda_{\Sigma_T}^{-1/2} \bar{V} \\ \bar{Y} = \bar{U}^\top \Lambda_{\Sigma_N}^{-1/2} \Phi^\top Y \Psi \Lambda_{\Sigma_T}^{-1/2} \bar{V}$$

Proof. : Taking the derivative of the cost function and setting the result to zero in a similar fashion to the previous proof results in the following expression for $F^\star$: □

$$\operatorname{vec}(F^\star) = \Big(S_T^\top \Sigma_T^{-1} S_T \otimes S_N^\top \Sigma_N^{-1} S_N + \gamma K^{-1} \otimes H^{-2}\Big)^{-1} \\ \big(S_T^\top \Sigma_T^{-1} \otimes S_N^\top \Sigma_N^{-1}\big) \operatorname{vec}(Y)$$

Substituting in the eigenvalue decompositions for the covariance matrices Σ_N and Σ_T gives

$$= \Big(S_T^\top \Psi \Lambda_{\Sigma_T}^{-1} \Psi^\top S_T \otimes S_N^\top \Phi \Lambda_{\Sigma_N}^{-1} \Phi^\top S_N + \gamma K^{-1} \otimes H^{-2}\Big)^{-1} \\ \big(S_T^\top \Psi \Lambda_{\Sigma_T}^{-1} \Psi^\top \otimes S_N^\top \Phi \Lambda_{\Sigma_N}^{-1} \Phi^\top\big) \operatorname{vec}(Y)$$

Applying identity 8 from table 1 and substituting the second set of eigenvalue decompositions gives

$$= \big(K S_T^\top \Psi \Lambda_{\Sigma_T}^{-1/2} \otimes H^2 \Phi \Lambda_{\Sigma_N}^{-1/2}\big) \times \\ \Big(\gamma I_{T'} \otimes I_{N'} + \bar{V} \bar{\Lambda}_K \bar{V}^\top \otimes \bar{U} \bar{\Lambda}_H \bar{U}^\top\Big)^{-1} \times \\ \big(\Lambda_{\Sigma_T}^{-1/2} \Psi^\top \otimes \Lambda_{\Sigma_N}^{-1/2} \Phi^\top\big) \operatorname{vec}(Y)$$

$$= \big(K S_T^\top \Psi \Lambda_{\Sigma_T}^{-1/2} \bar{V} \otimes H^2 \Phi \Lambda_{\Sigma_N}^{-1/2} \bar{U}\big) \times \\ \Big(\gamma I_{T'} \otimes I_{N'} + \bar{\Lambda}_K \otimes \bar{\Lambda}_H\Big)^{-1} \times \\ \big(\bar{V}^\top \Lambda_{\Sigma_T}^{-1/2} \Psi^\top \otimes \bar{U}^\top \Lambda_{\Sigma_N}^{-1/2} \Phi^\top\big) \operatorname{vec}(Y)$$

$$= \big(K S_T^\top \Psi \Lambda_{\Sigma_T}^{-1/2} \bar{V} \otimes H^2 \Phi \Lambda_{\Sigma_N}^{-1/2} \bar{U}\big) \times \\ \operatorname{diag}\big(\operatorname{vec}(J)\big) \operatorname{vec}(\bar{U}^\top \Lambda_{\Sigma_N}^{-1/2} \Phi^\top Y \Psi \Lambda_{\Sigma_T}^{-1/2} \bar{V})$$

$$= \big(C \otimes B\big)\, \operatorname{diag}\big(\operatorname{vec}(J)\big)\, \operatorname{vec}(\bar{Y})$$

Finally, applying identities 1 and 6 completes the proof.

Theorem II.3. *The cost function*

$$\xi(\theta) = N(T-1) \ln(1-\theta^2) \\ + \operatorname{tr}\big(E \Sigma_T^{-1}(\theta) E^\top \Sigma_N\big) + \frac{NT\alpha}{1-\theta^2}$$

is minimised by a value of θ that is the solution to the following cubic polynomial:

$$\frac{b}{2} + \big(N(T-1) - a - c - NT\alpha\big)\theta + \frac{b}{2}\theta^2 - N(T-1)\,\theta^3$$

where

$$a = \operatorname{tr}\big(E^\top \Sigma_N^{-1} E\big), \quad b = \operatorname{tr}\big(E^\top \Sigma_N^{-1} E B_1\big), \\ \text{and} \quad c = \operatorname{tr}\big(E^\top \Sigma_N^{-1} E B_2\big)$$

Proof. First expand Σ_T^{-1} in terms of θ.

$$\xi(\theta) = N(T-1) \ln(1-\theta^2) + \frac{\operatorname{tr}\big(E^\top \Sigma_N E\big) + NT\alpha}{1-\theta^2} \\ - \frac{\operatorname{tr}\big(E^\top \Sigma_N^{-1} E B_1\big)}{1-\theta^2}\theta + \frac{\operatorname{tr}\big(E^\top \Sigma_N^{-1} E B_2\big)}{1-\theta^2}\theta^2$$

Using the substitutions outlined previously, this simplifies.

$$\xi(\theta) = N(T-1) \ln(1-\theta^2) + \frac{a + NT\alpha - b\,\theta + c\,\theta^2}{1-\theta^2}$$

Taking the derivative with respect to θ gives

$$\frac{\partial \xi}{\partial \theta} = -\frac{2\theta N(T-1)}{1-\theta^2} + \frac{2\theta(a + NT\alpha + c) - b(\theta^2+1)}{(1-\theta^2)^2}$$

Setting this equal to zero and rearranging completes the proof.

□

Theorem II.4. *Given a fitted GLS KGR model, a lower bound for the marginal variance of the prediction uncertainty for the latent signal F is given by*

$$\Omega_F = \big(\tilde{U}^{-\top} \circ (H^2\tilde{U})\big)\, J\, \big(\tilde{V}^{-1} \circ (\tilde{V}^\top K)\big)$$

where $\circ$ is the Hadamard product,

$$S_T^\top \Sigma_T^{-1} S_T K = \tilde{V}\tilde{\Lambda}_K \tilde{V}^{-1}, \quad S_N^\top \Sigma_N^{-1} S_N H^2 = \tilde{U}\tilde{\Lambda}_H \tilde{U}^{-1}$$

and

$$J_{ij} = \frac{1}{\gamma + \tilde{\lambda}_i^K \tilde{\lambda}_j^H}$$

Proof. The Laplace approximation states that the posterior over the latent signal can be approximated by a Gaussian, with covariance given by the inverse Hessian (Fisher information matrix) of the posterior negative log likelihood.

$$\Sigma_F = \left[-\frac{\partial^2 \ln \pi(F|Y,X)}{\partial \text{vec}(F)_i \partial \text{vec}(F)_j}\bigg|_{F=F^\star} \right]^{-1}$$

To calculate this, first take the derivative of $-\ln \pi(F|Y,X)$ with respect to $\text{vec}(F)$.

$$\begin{aligned} -\frac{\partial \ln \pi(F|Y,X)}{\partial \text{vec}(F)} &= \big(S_T^\top \Sigma_T^{-1} S_T \otimes S_N^\top \Sigma_N^{-1} S_N \\ &\quad + \gamma K^{-1} \otimes H^{-2}\big)\text{vec}(F) - \text{vec}(S_N^\top \Sigma_N^{-1} Y \Sigma_T^{-1} S_T) \end{aligned}$$

Taking the Jacobian of this vector function of F gives the Hessian. From this, we can see tat Σ_F is given by

$$\Sigma_F = \big(S_T^\top \Sigma_T^{-1} S_T \otimes S_N^\top \Sigma_N^{-1} S_N + \gamma K^{-1} \otimes H^{-2}\big)^{-1}$$

First, factorise Σ_F as follows

$$= \big(K \otimes H^2\big)\Big(S_T^\top \Sigma_T^{-1} S_T K \otimes S_N^\top \Sigma_N^{-1} S_N H^2 + \gamma I_T \otimes I_N\Big)^{-1}$$

Then substitute in the eigenvalue decompositions as previously defined.

$$= \big(K \otimes H^2\big)\Big(\tilde{V}\tilde{\Lambda}_K \tilde{V}^{-1} \otimes \tilde{U}\tilde{\Lambda}_H \tilde{U}^{-1} + \gamma I_T \otimes I_N\Big)^{-1}$$

Next, factorise this expression further

$$\begin{aligned} &= \big(K \otimes H^2\big)\big(\tilde{V} \otimes \tilde{U}\big) \\ &\qquad \times \big(\tilde{\Lambda}_K \otimes \tilde{\Lambda}_H + \gamma I_T \otimes I_N\big)^{-1}\big(\tilde{V}^{-1} \otimes \tilde{U}^{-1}\big) \end{aligned}$$

Next, substitute in the definition of J.

$$= \big(K\tilde{V} \otimes H^2\tilde{U}\big)\, \text{diag}\big(\text{vec}(J)\big)\, \big(\tilde{V}^{-1} \otimes \tilde{U}^{-1}\big)$$

Since we are interested in the marginal variances, i.e. the diagonal of this matrix, we would like to avoid holding the full $(NT \times NT)$ matrix in memory. In order to achieve this, first consider the fact that the above matrix can be expressed as

$$\Sigma_F = \sum_{t=1}^{T} K\tilde{V}\delta_t\tilde{V}^{-1} \otimes H^2\tilde{U}\,\text{diag}(J_t)\,\tilde{U}^{-1}$$

where δ_t is an empty $T \times T$ matrix, except element (t,t), which is 1, and J_t is the tth column of J. This is effectively splitting each of the T operations involved in the Kronecker-product operation.

Now consider the diagonals of the above matrices. The diagonal of $K\tilde{V}\delta_t\tilde{V}^{-1}$ is the Hadamard product between the tth row of $\tilde{V}^{-1}$ and the tth column of $K\tilde{V}$. Denote this as $\mathbf{a}_t$.

$$\mathbf{a}_t = (\tilde{V}^{-\top})_t \circ (K\tilde{V})_t$$

The diagonal of the second matrix, $\mathbf{b}_t$, is

$$\mathbf{b}_t = \text{diag}^{-1}\big(H^2\tilde{U}\,\text{diag}(J_t)\,\tilde{U}^{-1}\big)$$

where $\text{diag}^{-1}(\cdot)$ means to take the diagonal of a matrix and turn it into a vector. Using identity 9 from table 1, we can rewrite $\mathbf{b}_t$ as

$$\mathbf{b}_t = \big(\tilde{U}^{-\top} \circ (H^2\tilde{U})\big) J_t$$

Therefore, Ω_F can be expressed as

$$\begin{aligned} \Omega_F &= \text{mat}\Big(\sum_{t=1}^{T} \mathbf{a}_t \otimes \mathbf{b}_t\Big) \\ &= \sum_{t=1}^{T} \mathbf{b}_t\, \mathbf{a}_t^\top \\ &= BA^\top \\ &= \big(\tilde{U}^{-\top} \circ (H^2\tilde{U})\big)\, J\, \big(\tilde{V}^{-1} \circ (\tilde{V}^\top K)\big) \end{aligned}$$

□

References

[1] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*, 2012.

[2] T. P. Minka, “Old and New Matrix Algebra Useful for Statistics,” Tech. Rep., 2001. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.115.5808