

Technical University of Crete

School of Electrical and Computer Engineering

SLIMNETS

Sparse Learning for Interacting Multiscale Networks

Annual Project Report

Reporting Year: 2026

Principal Investigator: Dionissios Hristopulos
Host Institution: Technical University of Crete
Funding Agency: Hellenic Foundation for Research and Innovation (HFRI)
Project Number: 25024



October 7, 2026

Contents

1	Work Package 1: Management and Dissemination (Integration)	3
1.1	Objectives	3
1.2	Progress During the Reporting Period	3
1.3	Deliverables	4
1.4	Deviations and Corrective Actions	5
2	Work Package 2: Boosting GPR	6
2.1	Objectives	6
2.2	Progress During the Reporting Period	6
2.2.1	Milestone MI.4: Review of Multi-task Covariance Kernels	6
2.2.2	Milestone MI.5: Initial Formulation of the Effective GPR Model	7
2.2.3	Karhunen–Loève Expansion of the Brownian Oscillator Kernel	8
2.3	Deliverables	9
2.4	Deviations and Corrective Actions	9
3	Work Package 3: Development of ST Local Interaction Machines (SLIM)	10
3.1	Objectives	10
3.2	Progress During the Reporting Period	10
3.3	Deliverables	11
3.4	Deviations and Corrective Actions	11
4	Work Package 4: SLIM and GP regression networks (Networks)	12
4.1	Objectives	12
4.2	Progress During the Reporting Period	12
4.3	Deliverables	12
4.4	Deviations and Corrective Actions	13
A	Convexity of SLI NLL	14
A.1	Gradient-based SLI energy function	14
A.2	SLI Precision Matrix Formulation	15
A.3	ML estimation	16
A.4	A simple and convex parametrization	17
A.4.1	Further simplification	17

Research Team

Internal Members

- **Dionissios Hristopoulos** — Professor, School of Electrical and Computer Engineering, Technical University of Crete
- **Athanasios Liavas** — Professor, School of Electrical and Computer Engineering, Technical University of Crete
- **Nikolaos Paraskakis** — PhD Candidate, School of Electrical and Computer Engineering, Technical University of Crete
- **Konstantinos Bekos** — PhD Candidate, School of Electrical and Computer Engineering, Technical University of Crete

External Collaborators - Advisory Board

- **Peter Atkinson** — Executive Dean, Faculty of Science and Technology, Lancaster University
- **Anastassia Baxevasi** — Associate Professor, Department of Mathematics and Statistics, University of Cyprus
- **Gerard Heuvelink** — Professor, Wageningen University; Senior Researcher, ISRIC–World Soil Information
- **Sandra de Iaco** — Professor, Department of Economic Science, University of Salento

External Scientific Collaborators

- **Sofia Nerantzaki** — Laboratory Teaching Staff, School of Mineral Resources Engineering, Technical University of Crete
- **Emmanouil Varouchakis** — Assistant Professor, School of Mineral Resources Engineering, Technical University of Crete

1 Work Package 1: Management and Dissemination (Integration)

1.1 Objectives

Coordination, data collection and pre-processing, central dissemination activities.

1.2 Progress During the Reporting Period

- An email letter was sent to the members of the Advisory Board on November 2, 2025, announcing that the project was officially launched on October 1, 2025. (**MI.1**).
- The dedicated project website was launched by November 2, 2025 (**MI.2, D1.1**)
- The purchase of equipment and software was completed (**MI.3**).
- Two new PhD candidates in the School of Electrical and Computer Engineering of the Technical University of Crete (TUC) *Nikolaos Paraskakis* and *Konstantinos Bekos* joined the project.
- Weekly meetings are held every Tuesday at 11AM in the meeting room of the School of Electrical and Computer Engineering. The typical duration of the meetings is about 1 hour. The meetings are regularly attended by the core SLIMNET members: D. Hristopulos (professor), A. Liavas (professor), N. Paraskakis (PhD candidate), K. Bekos (PhD candidate).

Some of the meetings have been attended by collaborators (Dr. Sofia Nerantzaki) and visiting faculty (Prof. Patrick Bogaert, Universite de Louvain, Belgium—in the period between February and May 2025).

- *Emmanouil Varouchakis* (Assistant Professor in the School of Mineral Resources Engineering, TUC) and *Sofia Nerantzaki* (Laboratory Teaching Staff in the School of Mineral Resources Engineering, TUC) joined the project in the role of collaborators.
- The main ideas and progress achieved by the project were presented at several international conferences:

1. “On Space-time Local Interaction Machines” was presented by Dionissios Hristopulos at the *19th International Joint Conference CFE-CMStatistics 2025*, held at Birkbeck, University of London, <https://www.slimnets.tuc.gr/en/new-s/item/presentation-titled-on-space-time-local-interaction-machi>

[nes-at-19th-international-joint-conference-cfe-cmstatistics-birkbeck-college-london.](#)

2. “Spatial and Spatio-Temporal Statistics” was presented by Dionissios Hristopulos at the *Joint Meeting SIS-FENStatS 2026*, hosted by the Department of Statistical Sciences at Sapienza University of Rome <https://www.slimnets.tuc.gr/en/news/item/presentation-at-sis-fenstats-2026>.
3. Three presentations were made by the team at the *SigmaPhi 2026 International Conference on Statistical Physics* held in Kolymbari, Crete. <https://www.slimnets.tuc.gr/en/news/item/our-team-at-sigmaphi-2026>. The titles of the presentations and the authors (presenting author in italics) are:
 - “Sparse Precision (Interaction) Kernels for Machine Learning”, *Dionissios Hristopulos*, Athanasios Liavas, Konstantinos Bekos, Nikolaos Paraskakis
 - “Karhunen-Loeve Expansion of the Brownian Oscillator Kernel,” *Nikolaos Paraskakis*, Dionissios Hristopulos
 - “Monthly Sunspot Number Forecasting using Gaussian Process Regression with Brownian Oscillator Kernel,” *Konstantinos Bekos*, Dionissios Hristopulos
4. “Recent Advances in Space-time Modeling with Stochastic Local Interaction Machines (SLIMs)”, *Dionissios Hristopulos*, Athanasios Liavas, Konstantinos Bekos, Nikolaos Paraskakis, presented at the *24th Annual Meeting of the International Association for Mathematical Geosciences (IAMG 2026)*, held in Montreal, Canada.

In addition, the following lecture was presented at the Technical University of Crete (it was also remotely accessible via Zoom). *EURECA-PRO: Invited Lecture* by Dionissios Hristopulos, “From Particles to Patterns: An Odyssey from Physics to Geostatistics and Beyond”, <https://www.slimnets.tuc.gr/en/news/item/eureca-pro-invited-lecture-by-dionissios-hristopulos-tuc>

Number	Milestone Name	Due Date (month)	Means of verification
MI.1	Inform Advisory Board regarding SLIMNETS start	M1	Letters sent by PI
MI.2	Launching dedicated SLIMNETS website	M2	Launch at GSLAB
MI.3	Finalize purchasing of equipment	M5	Invoices
MI.15	End-of-project workshop at TUC	M40	Abstracts, expense receipts, event program

Table 1.1: List of Milestones for WP1. Emphasized: **Completed**.

1.3 Deliverables

- **Deliverable D1.1: website launch** was completed. The website can be accessed at the address: <https://www.slimnets.tuc.gr/en/home>. It is regularly updated by team members, posting news and progress related to the project.

Number	Deliverable Name	Type	Level	Due Date (month)
D1.1	SLIMNETS web site launch	Other	PU	M2
D1.2	Annual project status report posted on SLIMNETS website	R	PU	M13
D1.3	Annual project status report posted on SLIMNETS website	R	PU	M25
D1.4	Annual project status report posted on SLIMNETS website	R	PU	M37
D1.5	Final project status report posted on SLIMNETS website	R	PU	M40
<i>D1.6</i>	<i>At least five presentations at Universities/Research Institutes</i>	DEC	PU	M40
D1.7	End-of-project workshop	Other	PU	M40

Table 1.2: List of Deliverables for WP1. PU: Public. R: Reports. DEC: Publications, conference presentations. Emphasized: **Completed**. Italicized: *Initiated/partly completed*.

- **Deliverable D1.2: Annual project report** for the first year of the project was completed.
- **Deliverable D1.6: At least five presentations at Universities/Research Institutes.** One presentation, titled *“From Particles to Patterns: An Odyssey from Physics to Geostatistics and Beyond”* was given by D. T. Hristopulos on June 9, 2026. <https://www.slimnets.tuc.gr/en/news/item/eureca-pro-invited-lecture-by-dionissios-hristopulos-tuc>

1.4 Deviations and Corrective Actions

There were no deviations from the plan.

2 Work Package 2: Boosting GPR

2.1 Objectives

Boosting GPR: Boltzmann-Gibbs Multi-Output Covariance Models and Effective Gaussian process regression (GPR).

(N.1) The formulation of matrix covariance kernels for vector-valued GPs based on the Boltzmann-Gibbs GPR framework will be pursued. The new kernels will overcome current limitations in the joint dependence structure that require either (i) identical correlation ranges for all the vector components or (ii) kernels derived from linear combination of sets of independent latent variables (linear co-regionalization).

(N.2) Fast GPR approximations will be developed using effective methods. The most computationally demanding step in GPR for large datasets is the inversion of the covariance kernel which scales as $O(N^3)$ with data size, thus limiting the practical usefulness to datasets with $N \sim O(10^3)$. We will employ ideas from the “effective medium” framework of statistical physics to derive very fast approximations for GPR prediction. We aim to achieve better computational time and predictive performance than the nearest neighbor Gaussian process which is based on the Vecchia approximation.

Number	Milestone Name	Due Date (month)	Means of verification
MI.4	Review of multi-task covariance kernels	M6	Short report
MI.5	Initial formulation of effective GPR model	M12	Matlab/Python code
MI.7	Initiate validation of effective GPR with large data	M20	Matlab/Python code
MI.8	Novel B-G multi-task covariance kernels	M26	Short report

Table 2.1: List of Milestones for WP2. Emphasized: **Completed**.

2.2 Progress During the Reporting Period

2.2.1 Milestone MI.4: Review of Multi-task Covariance Kernels

A review of matrix-valued covariance kernels for multi-output Gaussian processes was completed during the reporting period. The review covered separable and non-separable constructions, the intrinsic and linear models of coregionalization, process-convolution models, multivariate Matérn cross-covariances, conditional and graphical constructions, and non-stationary and space-time extensions. Particular attention was given to positive-definiteness requirements, component-specific correlation ranges and smoothness, the rep-

resentation of cross-component dependence, parameter identifiability, and computational cost.

The review first established the mathematical constraint shared by all such models. A valid matrix-valued kernel must be symmetric in its input–output arguments and generate a nonnegative-definite block covariance matrix for every finite set of input locations. Consequently, marginal covariance functions and pairwise cross-covariances cannot be specified independently. They must be chosen jointly. This validity requirement is the principal difficulty in designing kernels that are both flexible and interpretable.

Separable kernels, also known as intrinsic coregionalization models, are attractive because they combine a scalar input kernel with a positive-semidefinite output covariance matrix and can exploit Kronecker structure for co-located observations. Their main limitation is that all outputs share the same normalized input correlation shape, smoothness, and range. The linear model of coregionalization extends this construction through a sum of separable terms, thereby allowing several dependence scales and more flexible marginal correlations. However, its cross-dependence is still generated by linear mixing of latent processes, and the associated parameterizations may become large or weakly identifiable.

The review also examined alternatives that relax these restrictions. Process-convolution kernels provide a generative smoothing mechanism and can produce non-separable, output-specific covariance structures, but closed-form expressions and scalable inference are available only in selected cases. Multivariate Matérn models offer interpretable component-specific variance, range, and smoothness parameters. Their cross-parameters must nevertheless satisfy coupled validity conditions that become harder to enforce as the number of outputs increases. Conditional and graphical constructions can represent directed or sparse dependence, including asymmetric cross-covariances, whereas non-stationary and space–time extensions can model locally changing ranges, anisotropy, and cross-component relationships at the cost of additional parameters and data requirements.

From an inferential perspective, direct multi-output Gaussian-process calculations operate on an $np \times np$ covariance matrix for n observations of p outputs. Exploiting separability, matrix-normal structure, sparsity, or reduced-rank approximations is therefore essential when evaluating more flexible models. The review concludes that the kernel to be developed under Milestone MI.8 should permit distinct component-specific correlation ranges, represent cross-component dependence beyond simple latent linear mixing, make positive definiteness transparent, and retain enough structure for reliable parameter estimation and scalable prediction.

2.2.2 Milestone MI.5: Initial Formulation of the Effective GPR Model

Taking cues from mean-field theories of statistical physics, an innovative algorithm for fast GPR prediction has been developed.

- The algorithm has been implemented in Matlab code and applied to various synthetic data sets.
- It has also been applied to a real time series data set (monthly number of sunspots) and the results of the produced predictions were part of a poster presentation with title “Monthly sunspot number forecasting using Gaussian process regression with Brownian oscillator kernel” in the *International Conference on Statistical Physics* ($\Sigma\Phi$ 2026), held in Kolymbari, Crete, Greece.

- The computation time of the proposed algorithm is up to 10 times smaller than the time required by the exact GPR prediction and up to 20 times smaller than the nearest neighbor approximation demands.
- The predictive performance of the new algorithm is close to the performances of the exact GPR and the nearest neighbor approximation, although not better than them.

Beyond that, the algorithm is aimed to be improved and applied to larger data sets, in order to assess and validate its performance.

2.2.3 Karhunen–Loève Expansion of the Brownian Oscillator Kernel

In parallel with the work under Milestone MI.5, the Karhunen–Loève (KL) expansion of the Brownian oscillator (BO) covariance kernel was investigated as an efficient reduced-rank representation for Gaussian-process models. This work contributes directly to objective N.2 by replacing dense covariance calculations with a representation in terms of a limited number of orthogonal modes.

The principal developments achieved during the reporting period are as follows:

- The stationary BO covariance kernel was expressed in a unified parameterization covering the underdamped, critically damped, and overdamped regimes.
- The Fredholm eigenvalue problem for the BO kernel on a finite interval was reduced to a fourth-order homogeneous boundary-value problem. Semi-explicit eigenfunction branches and their boundary conditions were derived, with admissible eigenvalues determined by the singularity of a 4×4 boundary-condition matrix.
- A numerical BO–KL solver was developed using determinant-based root searches, singular-value checks, stabilized boundary-matrix representations, and singular-value-decomposition-based recovery of normalized eigenfunctions.
- The rank- M Mercer truncation was incorporated into GPR in KL-coefficient space. For $M \ll N$, the resulting formulation reduces the principal factorization cost from $O(N^3)$ for exact GPR to $O(NM^2 + M^3)$ and the principal storage requirement from $O(N^2)$ to $O(NM + M^2)$.
- Preliminary experiments on the monthly Niño 3.4 sea-surface-temperature index examined contiguous and scattered interpolation as well as forecasting. The results show accurate reconstruction in the interpolation tasks and the expected increase in error and uncertainty for extrapolation as the forecast horizon grows.

These developments establish the analytical and computational basis for using BO–KL modes in reduced-complexity interpolation, forecasting, simulation, and signal reconstruction.

Number	Deliverable Name	Type	Level	Due Date (month)
D2.1	At least one conference presentation on matrix B-G kernels	DEC	PU	M40
D2.2	At least one publication (journal or preprint) on multi-task kernels	DEC	PU	M40
D2.3	At least one conference presentation on effective GPR	DEC	PU	M40
D2.4	At least one publication (journal or preprint) on effective GPR	DEC	PU	M40

Table 2.2: List of Deliverables for WP2. PU: Public. R: Reports. DEC: Publications, conference presentations.

2.3 Deliverables

2.4 Deviations and Corrective Actions

There were no deviations from the plan.

3 Work Package 3: Development of ST Local Interaction Machines (SLIM)

3.1 Objectives

Formulation of novel space-time local interaction machines—implementing innovations N.3 and N.4 of Action 2.

(N.3): The development of novel sparse local interaction machines (SLIM) incorporating distance-dependent intensity factors, ST interaction operators, seasonal couplings, and non-Euclidean distances. The new SLIM will provide general and flexible regression tools for complex ST associations. Accurate and computationally efficient regression and simulation algorithms for gappy space-time datasets will be designed using SLIM.

(N.4): Generalization of the SLIM framework to multi-output and nonlinear SLIM for vector-valued Boltzmann-Gibbs processes. This innovation will advance the application of sparse SLIM models to regression problems that involve inter-dependent components. In addition, the introduction of nonlinear dependence via kernel functions will be pursued.

Number	Milestone Name	Due Date (month)	Means of verification
MI.6	SLIM estimation & regression algorithms report	M16	Short report
MI.9	SLIM estimation & regression algorithms code	M26	Matlab/Python code
MI.10	Initiate validation of SLIM methods with synthetic/real data	M30	Matlab/Python code
MI.11	SLIM simulation algorithm(s) for ST data	M32	Matlab/Python code
MI.12	Initiate assessment of SLIM performance	M32	Matlab/Python code

Table 3.1: List of Milestones for WP3.

3.2 Progress During the Reporting Period

The following items were investigated during the reporting period.

1. A direct connection was established between SLIMs and Boltzmann-Gibbs Gaussian processes which have an energy functional constructed by polynomials of the Laplacian operator.
2. For discretely-sampled space-time data, SLIMs can be constructed using a precision (inverse covariance) matrix that involves a polynomial of the space-time Laplacian defined on the respective space-time graph.

3. Suitable representations for the space-time graph Laplacian have been explored both on regular and irregular meshes.
4. Two forms were considered for the space-time Laplacian operator $L_{s,t}$: the first approach employs a space-time product, i.e., $L_{s,t} = L_s \otimes L_t$ where L_s, L_t are the spatial and temporal Laplacians, whereas the second approach involves a separable model, i.e., $L_{s,t} = L_s \otimes I_t + \alpha L_t \otimes I_s$ where I_s and I_t are spatial/temporal identity matrices and α is a temporal scaling factor.
5. It was rigorously proved that if c_0, c_1, c_2 are non-negative constants and $c_0 > 0$, and the graph Laplacian operator is also positive-definite, then precision operators of the form

$$Q = c_0 + c_1 L_{s,t} + c_2 L_{s,t}^2$$

are also positive definite. Furthermore, the likelihood of the SLIM model with energy function $H = \frac{1}{2} x^\top Q x$, where x is the vector of observations, is a convex function in the parameter space $\theta = (c_0, c_1, c_2)^\top$ (provided that any hyper-parameters involved in the definition of the space-time Laplacian, such as neighborhood-size controlling factors, remain fixed).

3.3 Deliverables

Number	Deliverable Name	Type	Level	Due Date (month)
D3.1	At least one conference presentation on multi-output SLIM models	DEC	PU	M32
D3.2	At least one publication (journal or preprint) on SLIM-models	DEC	PU	M40

Table 3.2: List of Deliverables for WP3. PU: Public. R: Reports. DEC: Publications, conference presentations.

3.4 Deviations and Corrective Actions

There were no deviations from the plan.

4 Work Package 4: SLIM and GP regression networks (Networks)

4.1 Objectives

Development of SLIM and GPR regression networks; implementing innovation N.5 of Action 2: SLIM Regression networks.

Gaussian Process Regression Networks (GPRNs) offer a principled approach to the analysis of simultaneous time series, combining the structural properties of Bayesian neural networks with the non-parametric flexibility of Gaussian Processes. We will extend this functionality to SLIM networks which will benefit from the sparse structure and fast computations that can be achieved by BGP components.

Number	Milestone Name	Due Date (month)	Means of verification
MI.13	Inference and regression methodology for effective GPR networks	M34	Matlab/Python code
MI.14	Inference and regression methodology for SLIM networks	M36	Short report

Table 4.1: List of Milestones for WP4.

4.2 Progress During the Reporting Period

WP4 is scheduled to begin in Month 14.

4.3 Deliverables

Number	Deliverable Name	Type	Level	Due Date (month)
D4.1	At least one conference presentation on GPR/SLIM networks	DEC	PU	M40
D4.2	1 publication (journal or preprint) on GPR/SLIM networks	DEC	PU	M40

Table 4.2: List of Deliverables for WP4. PU: Public. R: Reports. DEC: Publications, conference presentations.

4.4 Deviations and Corrective Actions

Not applicable since WP4 has not started yet.

A Convexity of SLI NLL

In this Appendix it is shown that the negative log-likelihood (NLL) of an SLI model with two terms is a convex function in the parameter space.

In the following,

1. $\mathbf{x}_s := (x_1, \dots, x_N)$ denotes the sample values at the locations of the sampling set $\mathbb{S}_N$.
2. $\mathbb{S}_N := \{\mathbf{s}_1, \dots, \mathbf{s}_N\}$, where $\mathbf{s}_i \in \mathcal{D} \subset \mathbb{R}^d$, for $i = 1, \dots, N$.
3. The prediction set involves the points $\mathbf{z}_p$, for $p = 1, \dots, P$, which correspond to the nodes of a prediction grid $\mathcal{P} \subset \mathbb{Z}^d$.

A.1 Gradient-based SLI energy function

The SLI model is defined in terms of a Gibbs–Boltzmann exponential joint density of the following form

$$f(\mathbf{x}_s) = \frac{1}{Z(\boldsymbol{\theta})} e^{-\mathcal{H}(\mathbf{x}_s; \boldsymbol{\theta})}, \quad (\text{A.1})$$

where $\boldsymbol{\theta}$, the SLI parameter vector to be estimated from the data, is defined below.

In the case of scattered data, a parsimonious SLI energy functional, that includes squared fluctuation and squared gradient terms, is given by

$$\mathcal{H}(\mathbf{x}_s; \boldsymbol{\theta}) := \frac{1}{2\lambda} \left(\frac{(\mathbf{x}_s - m_{\mathbf{x}} \mathbf{1})^T (\mathbf{x}_s - m_{\mathbf{x}} \mathbf{1})}{N} + c \mathcal{S}_1(\mathbf{x}_s; \mathbf{h}) \right), \quad (\text{A.2})$$

where

1. $m_{\mathbf{x}}$ is the mean value of $\mathbf{x}_s$,
2. $\lambda > 0$ is a positive scale coefficient proportional to the overall variability in the data,
3. $c > 0$ is a positive rigidity coefficient that weights the cost of gradients in the data,
4. $\mathcal{S}_1(\mathbf{x}_s; \mathbf{h})$ is the kernel–average square gradient, defined as

$$\begin{aligned} \mathcal{S}_1(\mathbf{x}_s; \mathbf{h}) &:= \langle (x_n - x_m)^2 \rangle_{\mathbf{h}} \\ &= \sum_{n=1}^N \sum_{m=1}^N w_{n,m}(\mathbf{h}) (x_n - x_m)^2, \end{aligned} \quad (\text{A.3})$$

where $\mathbf{h} := (h_1, \dots, h_N)$ is the vector of the local kernel bandwidths, defined as

$$h_n := \mu D_{n,[k]}(\mathbb{S}_N). \quad (\text{A.4})$$

$D_{n,[k]}(\mathbb{S}_N)$ is the distance between $\mathbf{s}_n$ and its k -th nearest neighbor, and μ is a positive bandwidth tuning parameter.

Quantities $w_{n,m}(\mathbf{h})$ are defined as

$$w_{n,m}(\mathbf{h}) := \frac{K(\mathbf{r}_{n,m}/h_n)}{\sum_{l=1}^N \sum_{q=1}^N K(\mathbf{r}_{l,q}/h_l)}, \text{ for } n, m = 1, \dots, N. \quad (\text{A.5})$$

We can write $w_{n,m}$ as a function of μ as follows:

$$w_{n,m}(\mu) := \frac{K(\mathbf{r}_{n,m}/(\mu D_{n,[k]}))}{\sum_{l=1}^N \sum_{q=1}^N K(\mathbf{r}_{l,q}/(\mu D_{l,[k]}))}, \text{ for } n, m = 1, \dots, N. \quad (\text{A.6})$$

5. For fixed k , $\boldsymbol{\theta}_k := (\mathbf{m}_x, \lambda, c, \mu)$ is the unknown parameter vector.

A.2 SLI Precision Matrix Formulation

The SLI energy function is a quadratic function of the sample values. Hence, it can be equivalently expressed in the following form that involves the precision matrix $\mathbf{J}(\boldsymbol{\theta}')$:

$$\mathcal{H}(\mathbf{x}_s; \boldsymbol{\theta}_k) = \frac{1}{2} (\mathbf{x}_s - \mathbf{m}_x \mathbf{1})^T \mathbf{J}(\boldsymbol{\theta}'_k) (\mathbf{x}_s - \mathbf{m}_x \mathbf{1}), \quad (\text{A.7})$$

where

$$\boldsymbol{\theta}'_k := (\lambda, c, \mu). \quad (\text{A.8})$$

The precision matrix has the following structure (note that $\mathbf{h}$ depends on k and μ):

$$\mathbf{J}(\boldsymbol{\theta}'_k) := \frac{1}{\lambda} \left(\frac{1}{N} \mathbf{I}_N + c \mathbf{J}_1(\mathbf{h}) \right). \quad (\text{A.9})$$

In the above:

1. $\mathbf{I}_N$ is the $N \times N$ identity matrix, which results from the energy of the squared fluctuations,
2. and $\mathbf{J}_1(\mathbf{h})$ is the gradient precision submatrix which results from the sum of the squared differences, with

$$(\mathbf{J}_1(\mathbf{h}))_{n,m} = -w_{n,m}(\mathbf{h}) - w_{m,n}(\mathbf{h}) + \delta_{n,m} \sum_{l=1}^N (w_{n,l}(\mathbf{h}) + w_{l,n}(\mathbf{h})). \quad (\text{A.10})$$

In terms of μ , $(\mathbf{J}_1(\mathbf{h}))_{n,m}$ is expressed as

$$(\mathbf{J}_1(\mu))_{n,m} = -w_{n,m}(\mu) - w_{m,n}(\mu) + \delta_{n,m} \sum_{l=1}^N (w_{n,l}(\mu) + w_{l,n}(\mu)). \quad (\text{A.11})$$

Finally, $\mathbf{J}(\cdot)$ can be expressed as

$$\mathbf{J}(\boldsymbol{\theta}'_k) = \mathbf{J}(\lambda, c, \mu) = \frac{1}{\lambda} \left(\frac{1}{N} \mathbf{I}_N + c \mathbf{J}_1(\mu) \right). \quad (\text{A.12})$$

A.3 ML estimation

The SLI negative log-likelihood is given by

$$\begin{aligned} -\ln \mathcal{L}(\boldsymbol{\theta}_k; \mathbf{x}_s) &= \mathcal{H}(\mathbf{x}_s; \boldsymbol{\theta}_k) + \ln Z(\boldsymbol{\theta}_k) \\ &= \frac{1}{2} (\mathbf{x}_s - m_{\mathbf{x}} \mathbf{1})^T \mathbf{J}(\boldsymbol{\theta}'_k) (\mathbf{x}_s - m_{\mathbf{x}} \mathbf{1}) + \ln Z(\boldsymbol{\theta}_k). \end{aligned} \quad (\text{A.13})$$

The partition function for the SLI pdf is given by

$$Z(\boldsymbol{\theta}_k) = (2\pi)^{N/2} [\det \mathbf{J}(\boldsymbol{\theta}'_k)]^{-1/2} = (2\pi\lambda)^{N/2} [\det \tilde{\mathbf{J}}(\boldsymbol{\theta}'_{k,-\lambda})]^{-1/2}, \quad (\text{A.14})$$

where

$$\boldsymbol{\theta}'_{k,-\lambda} := (c, \mu), \quad (\text{A.15})$$

and $\tilde{\mathbf{J}}(\boldsymbol{\theta}'_{k,-\lambda}) = \lambda \mathbf{J}(\boldsymbol{\theta}'_k)$ is the *scale-free precision matrix*, which is *independent* of λ . In detail, we have

$$\tilde{\mathbf{J}}(\boldsymbol{\theta}'_{k,-\lambda}) = \tilde{\mathbf{J}}(c, \mu) = \frac{1}{N} \mathbf{I}_N + (c\lambda) \mathbf{J}_1(\mu) \stackrel{(a)}{=} \frac{1}{N} \mathbf{I}_N + c \mathbf{J}_1(\mu), \quad (\text{A.16})$$

where at point (a) we call, again, the variable c .

Thus,

$$-\ln \mathcal{L}(\boldsymbol{\theta}_k; \mathbf{x}_s) = \frac{1}{\lambda} \tilde{\mathcal{H}}(\mathbf{x}_s; \boldsymbol{\theta}_{k,-\lambda}) + \frac{N}{2} \ln 2\pi\lambda - \frac{1}{2} \ln \det \tilde{\mathbf{J}}(\boldsymbol{\theta}'_{k,-\lambda}). \quad (\text{A.17})$$

where $\tilde{\mathcal{H}}(\mathbf{x}_s; \boldsymbol{\theta}_{k,-\lambda})$ is the energy function evaluated at $\lambda = 1$, that is,

$$\tilde{\mathcal{H}}(\mathbf{x}_s; \boldsymbol{\theta}_{k,-\lambda}) = \frac{1}{2} (\mathbf{x}_s - m_{\mathbf{x}} \mathbf{1})^T \left(\frac{1}{N} \mathbf{I}_N + c \mathbf{J}_1(\mu) \right) (\mathbf{x}_s - m_{\mathbf{x}} \mathbf{1}), \quad (\text{A.18})$$

Equation (A.17) can be explicitly minimized, with respect to λ , yielding

$$\lambda^* = \frac{2\tilde{\mathcal{H}}(\mathbf{x}_s; \boldsymbol{\theta}_{k,-\lambda})}{N}. \quad (\text{A.19})$$

If we use this expression in (A.17), we obtain

$$-\ln \mathcal{L}(\boldsymbol{\theta}_k; \mathbf{x}_s) = \frac{N}{2} \ln 2\pi e + \frac{N}{2} \ln \left(\frac{2\tilde{\mathcal{H}}(\mathbf{x}_s; \boldsymbol{\theta}_{k,-\lambda})}{N} \right) - \frac{1}{2} \ln \det \tilde{\mathbf{J}}(\boldsymbol{\theta}'_{k,-\lambda}). \quad (\text{A.20})$$

This can be written as

$$\begin{aligned} F(m_{\mathbf{x}}, c, \mu) &= \frac{N}{2} \ln \left[(\mathbf{x}_s - m_{\mathbf{x}} \mathbf{1})^T \left(\frac{1}{N} \mathbf{I}_N + c \mathbf{J}_1(\mu) \right) (\mathbf{x}_s - m_{\mathbf{x}} \mathbf{1}) \right] \\ &\quad - \frac{1}{2} \ln \det \left[\frac{1}{N} \mathbf{I}_N + c \mathbf{J}_1(\mu) \right] \\ &=: F_1(m_{\mathbf{x}}, c, \mu) + F_2(c, \mu). \end{aligned} \quad (\text{A.21})$$

where

$$F_1(m_{\mathbf{x}}, c, \mu) := \frac{N}{2} \ln \left[(\mathbf{x}_s - m_{\mathbf{x}} \mathbf{1})^T \left(\frac{1}{N} \mathbf{I}_N + c \mathbf{J}_1(\mu) \right) (\mathbf{x}_s - m_{\mathbf{x}} \mathbf{1}) \right] \quad (\text{A.22})$$

and

$$F_2(c, \mu) := -\frac{1}{2} \ln \det \left[\frac{1}{N} \mathbf{I}_N + c \mathbf{J}_1(\mu) \right] \quad (\text{A.23})$$

We must minimize function $F(m_{\mathbf{x}}, c, \mu)$.

A.4 A simple and convex parametrization

We consider an SLI model with precision matrix defined by

$$\mathbf{J}(\alpha, c_1) = \alpha \mathbf{I} + c_1 \mathbf{J}_1 \in \mathbb{R}^{n \times n}, \quad (\text{A.24})$$

where $\alpha = c_0/N > 0$ and $\mathbf{J}_1$ is the graph Laplacian. We define

$$\mathbf{x}' := \mathbf{x} - m_{\mathbf{x}} \mathbf{1}. \quad (\text{A.25})$$

Then, the Negative Log-Likelihood function is given by

$$\text{NLL}(\alpha, c_1) = \frac{1}{2} \mathbf{x}'^T \mathbf{J}(\alpha, c_1) \mathbf{x}' - \frac{1}{2} \log \det \mathbf{J}(\alpha, c_1). \quad (\text{A.26})$$

The function NLL is a convex function of α and c_1 . A proof follows:

1. $\mathbf{J}$ is a linear function of α and c_1 , thus, the first term is convex;
2. function $-\frac{1}{2} \log \det(\cdot)$ is convex, and its synthesis with a linear function remains convex;
3. the sum of two convex functions is convex.

The dependence of $\mathbf{J}_1$ on μ is non-convex. Thus, it seems reasonable to fix μ and optimize with respect to α and c_1 , and repeat the process over a certain range of μ 's.

A.4.1 Further simplification

Recall that $\mathbf{J}(\alpha, c_1) = \alpha \mathbf{I} + c_1 \mathbf{J}_1 \in \mathbb{R}^{n \times n}$.

We simplify function NLL as much as possible.

$$\begin{aligned} \text{NLL}(\alpha, c_1) &= \frac{1}{2} \mathbf{x}'^T \mathbf{J}(\alpha, c_1) \mathbf{x}' - \frac{1}{2} \log \det \mathbf{J}(\alpha, c_1) \\ &= \frac{1}{2} (\alpha \|\mathbf{x}'\|^2 + c_1 \mathbf{x}'^T \mathbf{J}_1 \mathbf{x}') - \frac{1}{2} \log \det \mathbf{J}(\alpha, c_1). \end{aligned} \quad (\text{A.27})$$

Further,

$$\det \mathbf{J}(\alpha, c_1) = \prod_{i=1}^n (\alpha + c_1 \lambda_i(\mathbf{J}_1)), \quad (\text{A.28})$$

where $\lambda_i(\mathbf{J}_1)$ is the i -th eigenvalue of $\mathbf{J}_1$. Thus,

$$\log \det \mathbf{J}(\alpha, c_1) = \log \left(\prod_{i=1}^n (\alpha + c_1 \lambda_i(\mathbf{J}_1)) \right) = \sum_{i=1}^n \log (\alpha + c_1 \lambda_i(\mathbf{J}_1)). \quad (\text{A.29})$$

We define $z = \|\mathbf{x}'\|_2$ and $w = \mathbf{x}'^T \mathbf{J}_1 \mathbf{x}'$.

Then, the simplest expression of the NLL function is as follows:

$$\text{NLL}(\alpha, c_1) = \frac{\alpha}{2} z^2 + \frac{c_1}{2} w - \frac{1}{2} \sum_{i=1}^n \log (\alpha + c_1 \lambda_i(\mathbf{J}_1)). \quad (\text{A.30})$$

In Fig. A.1, we plot the level sets of the cost function $\text{NLL}(\alpha, c_1)$ for the dataset included in `example4`.

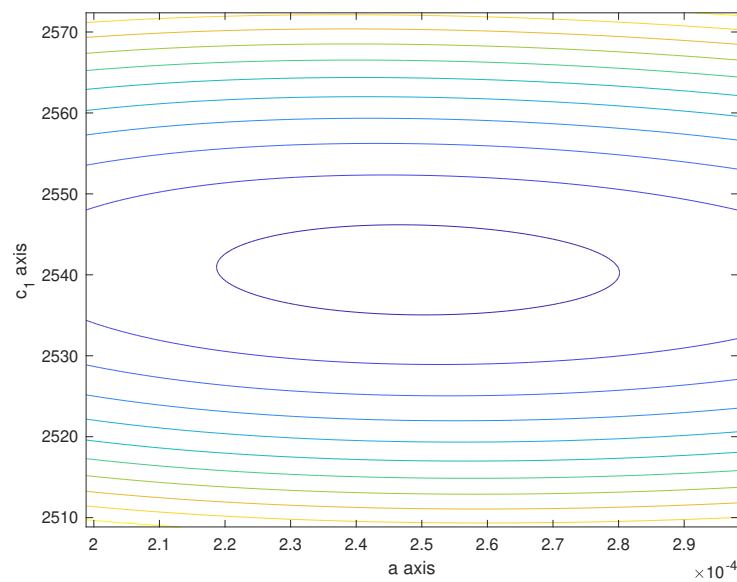


Figure A.1: Level set of simplified form of cost function NLL.