

Inhalt

| | |
|--|----|
| Editorial | 2 |
| Iterationsschleife | 5 |
| Datafold: representation learning and time series forecasting | 7 |
| IN.TUM goes Heilbronn | 15 |
| Supporting Shared-Memory Parallel Sparse BLAS from C/C++ up to Interpreted Languages | 17 |
| Erster NHR-Rechner für Erlangen | 21 |
| NHR Perflab Seminar | 23 |
| Upcoming KONWIHR: Workshop October 11, 2021 | 24 |
| Notiz*Notiz*Notiz | 25 |

Quartl* - Impressum

Herausgeber:

Prof. Dr. A. Bode, Prof. Dr. H.-J. Bungartz, Prof. Dr. U. Rüde

Redaktion:

S. Herrmann, S. Seckler, Dr. S. Zimmer

Technische Universität München, Fakultät für Informatik

Boltzmannstr. 3, 85748 Garching b. München

Tel./Fax: ++49-89-289 18611/18607

e-mail: hermasa@in.tum.de,

<https://www.in.tum.de/index.php?id=5353>

Redaktionsschluss für die nächste Ausgabe: **01.12.2021**

* Quartl: früheres bayerisches Flüssigkeitsmaß,

→ das Quart: 1/4 Kanne = 0.27 l

(Brockhaus Enzyklopädie 1972)

Das Quartl erhalten Sie online unter <https://www.in.tum.de/index.php?id=5353>



Das Quartl ist das offizielle Mitteilungsblatt des *Kompetenznetzwerks für Technisch-Wissenschaftliches Hoch- und Höchstleistungsrechnen in Bayern* (KONWIHR) und der *Bavarian Graduate School of Computational Engineering* (BGCE)

Editorial

Seit 1984 gibt es sie, die Ferienakademie, und Leser:innen des Quartls sind natürlich bestens mit ihr vertraut. Zunächst als Veranstaltung der TUM ins Leben gerufen, gesellte sich kurz darauf die FAU Erlangen-Nürnberg als Mitveranstalter dazu, und im Jahre 2002 komplettierte schließlich die Universität Stuttgart das Trio. Jedes Jahr setzte sich seither zu Beginn des Oktoberfestes die Karawane in Bewegung gen Samtal. Jedes Jahr? Falsch – im vergangenen Jahr sorgte ein kleines Virus tatsächlich für einen Herbst ohne Ferienakademie. Zwar waren die Zahlen im letzten Sommer besser als jetzt, aber es gab ja noch keinen Impfstoff, die zweite Welle drohte (und kam dann ja auch mit voller Wucht über uns), und die Erfahrungen im Umgang mit einer Pandemie waren noch limitiert. So entschied man sich Anfang Juni 2020 schweren Herzens für die erste Absage – nach Ablauf der Bewerbungsfrist, aber vor der Auswahl der Teilnehmer:innen. Zwar gab es einen dreitägigen Kurzbesuch eines kleinen Kern-Teams im Samtal (ein kurzes Hallo bei den Gasthöfen, ein kurzer Check, ob Berge und Wanderwege noch da sind, sowie der eine oder andere Kaiserschmarr'n), aber das war keine Ferienakademie.

Neue Chance, neues Glück. Allen Entwicklungen des letzten Winters zum Trotz gingen alle entschlossen die Planungen für 2021 an: Festlegung der Kurse und Dozent:innen, Druck der Plakate, Werbung in den Lehrveranstaltungen, und schließlich im Frühjahr dann die Bewerbungsphase. Wir waren uns zunächst nicht sicher, wie sich die Studis verhalten würden: Würden Unsicherheit und Zurückhaltung dominieren, oder würde die Sehnsucht nach Normalität obsiegen? Nun, Letzteres war der Fall, die Anzahl der Bewerbungen erreichte ein Allzeit-Hoch – was unsere Motivation, es dieses Jahr aber hinzukriegen, natürlich stark beförderte.

Kernstück für eine solche Veranstaltung ist fraglos das Hygienekonzept. Und dieses geht über AHA, Maske und 3G weit hinaus. Schließlich sind gut 200 Leute nach Südtirol zu transportieren, in Bussen, mit unterschiedlichen Regelungen hierfür in Deutschland, Österreich und Italien. Die Unterbringung in den Gasthöfen muss in Zwei- und Dreibettzimmern erfolgen können.

Die zur Verfügung stehenden Flächen für Mahlzeiten und Kursstutzungen sind limitiert. Und natürlich: Was macht man im Fall der Fälle, wie organisiert man Quarantäne? Wer kommt dafür ggf. auf? Und wie regelt man die Stornofrage? Denn was können italienische Hotels dafür, wenn eine deutsche Uni auf die Notbremse tritt? Doch alle Fragen konnten geklärt werden – was rückblickend fast wie ein Wunder klingt. Das Hygienekonzept iterierte und iterierte und fand dann doch zur Konvergenz. Früh übrigens unsere Festlegung auf Impfpflicht – kein Durchimpfen, keine Ferienakademie. Auch das haben wir natürlich rechtlich prüfen lassen, und wir bekamen grünes Licht von den Experten (und ernteten keinerlei Protest von Seiten der Teilnehmenden). Dass ich dieses unsägliche Rum-Geeiere um eine generelle Impfpflicht (nur für alle, die können, selbstverständlich) eh nicht nachvollziehen kann, ist ein anderes Thema und würde ein eigenes Editorial füllen (und tut das vielleicht auch noch ...).

Am 8. September kam dann die finale Stellungnahme des Corona-Krisenstabs der TUM (mit dem wir immer die ersten Abstimmungen durchführen) zu unserem finalen Hygienekonzept. Und was da stand, liest man immer wieder so gerne, dass ich es hier zitieren möchte:

„Vielen Dank für die Zusendung des sehr vorbildlichen Infektionsschutzkonzepts. Es regelt aus meiner Sicht alle Punkte, die nach den in Bayern geltenden Regeln einzuhalten wären. Auch ohne die italienischen Regelungen im Detail zu kennen, sehe ich beim Lesen förmlich die Hervorhebung 'abweichender Standard Italien'. Daher gehe ich davon aus, dass Sie sich intensiv damit beschäftigt haben und auch diese Regelungen so implementiert haben, dass das Konzept jeglicher Überprüfung (wenn es eine gäbe...) standhalten würde. Die von uns angeregten Änderungen hinsichtlich des Konzepts bei Verdachtsfällen nach (a-)symptomatischer Schnelltestung und zur Lüftung sehe ich im finalen Konzept als bestmöglich gelöst an. Der zusätzliche Test vor Antritt (trotz erfolgter Impfung) ist für ein derartiges Format eine sehr sinnvolle Ergänzung, um asymptotische Impfdurchbrüche auszuschließen. Da die Ferienakademie noch im Zeitraum der kostenlosen Bürgertests stattfindet, ist dies auch sicherlich jedem zuzumuten.“

„Wow!“, so mein spontaner Gedanke – den unser TUM-Beauftragter und TUM-Vizepräsident Gerhard Müller noch wie folgt ausformulierte: „*Von ca. 10% eines solchen Feedbacks (mit die kritischste Einheit der TUM attestiert ein fulminantes Ergebnis) träume ich seit Beginn meiner Amtszeit*“.

Die Anerkennung für unser Hygienekonzept kommt nicht von ungefähr – da stecken viel Herzblut und Arbeit dahinter. An dieser Stelle daher ein besonders dickes Dankeschön an Tobias Neckel und Friedrich Menhorn, die dahinter stehenden Überlegungen angestellt, die zahlreichen Versionen verfasst, mit drei universitären Krisenstäben, der Südtiroler Gesundheitsbehörde, den Gasthöfen, dem Busunternehmen, dem Testzentrum auf dem Campus Garching (das vor der Busabfahrt am Sonntag nochmals eine Komplettestung vornimmt) und vielleicht sogar mit den Kastelruther Spatzen iteriert und trotzdem den Überblick bewahrt haben!

Ich schreibe dieses Editorial exakt eine Woche vor dem geplanten Start der diesjährigen Ferienakademie, und momentan sieht es tatsächlich so aus, dass wir am nächsten Sonntag unser Abenteuer beginnen können. Drücken wir die Daumen und hoffen wir, dass sich der ganze Aufwand auch gelohnt hat.

Die gesamte Quartl-Redaktion wünscht Ihnen allen, liebe Leserinnen und Leser, einen guten Start ins neue Studienjahr, das uns hoffentlich mehr Präsenz und mehr Normalität bringen möge als das letzte. Bleiben Sie gesund und haben Sie viel Spaß mit der neuesten (der 99.) Ausgabe Ihres Quartls!

Hans-Joachim Bungartz.

* Notiz * Notiz * Notiz *

Termine 2021 (falls Corona es erlaubt)

- **Upcoming SIAM Conferences & Deadline**
<https://www.siam.org/conferences/calendar>
- **Supercomputing 2021:**
The International Conference for High Performance Computing, Networking, Storage and Analysis (SC21) – SC 21 in St. Louis, USA: 14.11.-19.11.2021 <https://sc21.supercomputing.org/>

Upcoming KONWIHR: Workshop October 11, 2021



KONWIHR, the competence network for scientific high-performance computing in Bavaria¹³, organizes regular welcoming workshops for its projects. We invite the following recent projects to participate in our workshop on October 11, 2021 between 15:00-16:30:

- Optimisation of SeisSol for Large Scale Simulations of Induced Earthquakes
– Prof. Michael Bader (TUM SCCS)
- Fast and scalable finite element algorithms for coupled multiphysics problems and non-matching grids
– Dr. Martin Kronbichler, Prof. Wolfgang Wall (TUM LNM)
- Implementation of hyper-dimensional thermodynamic integration for parallel HPC molecular dynamics simulation using LAMMPS
– Prof. Dirk Zahn (FAU CCC)

You can find more details about these projects at

<https://www.konwihhr.de/konwihhr-projects/>

Join us to learn about these projects and find out about funding and HPC competence dissemination opportunities via KONWIHR and the recently established national HPC centers (NHR Alliance). The meeting will be held online¹⁴ and every Quartl reader is very welcome to join and participate in the discussions.

Gerasimos Chourdakis

¹³KONWIHR website: konwihhr.de. Contact KONWIHR at info@konwihhr.de.

¹⁴KONWIHR Workshops BBB: <https://bbb.in.tum.de/ger-ma-j-a6a>

Iterationsschleife

N=40

07. September 2021

Ende des 18. Jahrhunderts schreibt Hegel^a einen kleinen Text. Bekannt wurde er als „ältestes Systemprogramm des deutschen Idealismus“^b. Ob Hegel auch der Autor des Textes ist oder diesen nur abgeschrieben hat ist umstritten. Herfried Münkler^c spricht davon, dass die drei „Tübinger Stifter“ im Verdacht stehen, die Autoren zu sein^d. Hegel, Schelling^e oder Hölderlin^f kommen also als Autoren wohl auch in Frage. Sicher ist, dass es sich um Hegels Handschrift handelt. Ob er den Text jedoch abgeschrieben oder selber verfasst hat, bleibt weiterhin unklar, denn außer dem Vornamen „Friedrich“ hatten die drei Verdächtigen wohl auch noch einige gedankliche Gemeinsamkeiten.

Für Literaturwissenschaftler:innen ist der Text Labsal, denn – obwohl sehr kurz auf zwei Seiten – er kommt zur selbstsicheren Erkenntnis „die *dichtkunst allein wird alle übrigen Wissenschaften u[nd] Künste überleben*“.^g Eine gewagte Aussage schon in Bezug auf die Kunst. Malerei und Bildhauerei sind weitaus älter als die schriftliche Literatur und könnten aus dem höheren Alter womöglich auch die längere Lebensdauer ableiten. Andererseits bliebe zu argumentieren, dass wer früh beginnt womöglich auch früh altert. Das gemalte Bild als Kunstwerk ist bekanntlich seit 150 Jahren in der Krise zwischen Abstraktion und Konkretem, unter Druck gesetzt durch die Fotografie und aufgelöst durch die multimedialen Werke der digitalen Kunst^h.

^aGeorg Wilhelm Friedrich Hegel, Deutscher Philosoph, 27.08.1770 – 14.11.1831

^bChristoph Jamme und Helmut Schneider, Mythologie der Vernunft, suhrkamp, 1984

^cHerfried Münkler, Deutscher Politikwissenschaftler, geb. 1951

^dHerfried Münkler, Marx – Wagner – Nietzsche, rowohlt, 2021, p 66

^eFriedrich Wilhelm Joseph Schelling, Deutscher Philosoph, 27. Januar 1775 – 20. August 1854

^fJohann Christian Friedrich Hölderlin, Deutscher Dichter, 20. März 1770 – 7. Juni 1834

^gJamme & Schneider, 1984, p 13

^hEs lohnt ein Besuch im Zentrum für Kunst und Medien (ZKM) in Karlsruhe, <https://zkm.de>, der Österreicher Peter Weibel (geb. 1944 in Odessa) präsentiert dort nicht nur eine der spannendsten Sammlungen weltweit sondern auch eine media art in progress um die Karlsruhe international benedict wird.

Spannender ist aber eine andere zentrale Stelle des „Systemprogramms“, die sich mit dem Thema des Staates beschäftigt. *„die Idee der Menschheit voran – will ich zeigen, daß es keine Idee vom Staat gibt, weil der Staat etwas mechanisches ist, so wenig als es eine Idee von einer Maschine gibt. Nur was Gegenstand der Freiheit ist, heist Idee. Wir müssen also auch über den Staat hinaus! – Den jeder Staat muß freie Menschen als mechanisches Räderwerk behandeln; und] das soll er nicht; also soll er aufhören.“*^a

Das nun ist ein erstaunlicher Standpunkt. Man muss die Stelle logischerweise im Kontext der Zeit lesen. Der moderne Staat wie wir ihn kennen ist im besten Fall im Entstehen begriffen. Der Autor / Die Autoren können sich also nur auf eine Entwicklung beziehen, die die Herrschaft von der Einzelgewalt des Fürsten auf formale Prozesse und entstehende staatliche Bürokratien zu übertragen beginnt. Es heißt schließlich auch weiter *„Zugleich will ich hier die] Principien für eine Geschichte der Menschheit niederlegen, und] das ganze elende Menschenwerk von Staat, Verfaßung, Regierung, Gesetzgebung – bis auf die Haut entblösen.“*^b

Starke Worte, die irritieren und trotzdem so vertraut klingen. Mehr dazu in Iterationsschleife 41.

M. Resch

^aJamme & Schneider, 1984, pp 11-12

^bJamme & Schneider, 1984, p 12

NHR Perflab Seminar



Ein erklärtes Ziel des NHR-Verbundes ist der intensive Austausch über Ländergrenzen und Universitäten hinweg zu allen Themen, die den Betrieb und die Nutzung von Hochleistungsrechnern betreffen. Dazu gehören auch Forschungsergebnisse, „Best Practices“, Erfahrungen mit neuen Architekturen und Algorithmen, Softwarewerkzeuge, und eben alles, was über das eigene Zentrum und seine PIs hinaus interessant ist.

Deswegen haben sich die NHR-Zentren der FAU Erlangen-Nürnberg, der RWTH Aachen, des Zuse-Instituts Berlin und der Universität Paderborn zusammengetan, um eine regelmäßige Vortragsreihe zu organisieren, in der es sowohl interne (mit beschränktem Personenkreis) als auch öffentliche Beiträge gibt. Unter <https://hpc.fau.de/research/nhr-perflab-seminar-series/> finden sich Links auf vergangene und zukünftige Events dieser Reihe und – falls möglich – Vortragsfolien bzw. Videoaufzeichnungen.

Wer über anstehende Veranstaltungen informiert werden möchte, kann sich in einen Mailverteiler eintragen (der Link befindet sich auf der o. a. Seite). Vorschläge für zukünftige Beiträge werden selbstverständlich gerne entgegengenommen.

Georg Hager

Darüber hinaus werden zwei luftgekühlte GPGPU-Partitionen auf die Bedürfnisse der Anwender von Moleküldynamik- (MD) und KI-Applikationen zugeschnitten. Ein Teil der Rechenressourcen ist Tier-3-Kunden aus der FAU und aus Nordbayern gewidmet und wird daher mit einer niedrigen Einstiegsschwelle verfügbar sein. Der Großteil wird über ein wissenschaftliches und technisches Begutachtungsverfahren zugänglich gemacht, dessen Struktur derzeit von den NHR-Zentren finalisiert wird.

Georg Hager

Datafold: representation learning and time series forecasting

In “traditional” scientific computing, a typical workflow is to define a problem, design a numerical algorithm together with an implementation that can solve the problem, and finally execute everything to obtain a solution. The algorithms typically act iteratively and with a certain accuracy, and require some form of boundary or initial data. With the amount and complexity of “initial data” ever increasing, the field of machine learning has been deviating from this workflow to some degree. To obtain solutions to a problem, a typical workflow consists of data cleaning and preprocessing, the definition of an expressive class of models, and then iterative “training” with changes to many parameters with the goal of minimizing a problem-specific cost function. Finally, the sufficiently converged model is used as a solution to the initial problem. The class of expressive models nowadays often includes various types of neural networks.

The two approaches to numerical problem solving are not mutually exclusive. One quite principled and mathematically well-founded way to perform “scientific machine learning” is to design algorithms and their implementations for the numerical approximation of linear operators that act on functions on the data. Interesting and useful linear operators for data analysis and time-series forecasting are Laplace-Beltrami and Koopman operators.

At the chair for Scientific Computing in Computer Science of Hans-Joachim Bungartz (TUM), and in collaboration with the group of Gerta Köster at the Munich University of Applied Sciences (Hochschule München), we are developing the software datafold for representation and manifold learning through approximation of the Laplace operator. The software also allows researchers to model time series data on the learned geometry through numerical approximation of the Koopman operator. Datafold is available online as an open-source software package [1] and also briefly described in our recent paper [2].

More specifically, datafold is a Python package that provides data-driven models for point clouds to find an explicit mani-fold parametrization and to identify non-linear dynamical systems on these manifolds. The explicit data manifold treatment allows researchers to include prior knowledge of a system and its problem-specific domain. This can be proximity information between points in the dataset, or functions defined on the state space manifold of a dynamical system, such as (partially) known governing equation terms.

Representation Learning with the Laplace-Beltrami Operator

Expanding datasets create challenges throughout the analysis workflow, from extracting and processing to interpreting the data. This includes the fact that new data does not always provide completely new and uncorrelated information to existing data. One way to extract the essential information is to understand and parametrize the intrinsic data geometry. An intrinsic geometry is what most data-driven models assume implicitly or explicitly in the available data, and successful machine learning algorithms adapt to this underlying structure for tasks like regression or classification. This geometry is often of much lower dimension than the ambient data space, and finding a suitable set of coordinates can reduce the complexity of the dataset (see Figure 1). We refer to this geometric structure encoded in the data as a “manifold”.

In mathematical terms, a manifold is a topological space that is locally homeomorphic to Euclidean space. Our datafold software aims to construct a global parametrization (embedding) of this manifold, in a space of much lower dimension than the original ambient space. The manifold hypothesis states that such manifolds can approximate the geometry of many real observations and processes well, including the state spaces of time-dependent systems. In a machine learning context, this is also referred to as “representation learning”, “non-linear unsupervised learning”, or simply “manifold learning”. In

Erster NHR-Rechner für Erlangen



Ende des Jahres 2020 hat die Friedrich-Alexander-Universität Erlangen-Nürnberg ihr erstes NHR-System ausgeschrieben. Mit einem Gesamtvolumen von 10 Millionen € ist dies der größte Parallelrechner, der jemals an der FAU installiert wurde. Nach einem harten Wettbewerb ging die MEGWARE Computer Vertrieb und Service GmbH aus Chemnitz als Sieger hervor.



Abbildung 1: Struktur des NHR-Systems an der FAU.

Im Zeitrahmen Q4/21 bis Q1/22 wird nun ein System mit drei Partitionen aufgebaut: Ein wassergekühltes Multicore-Cluster auf Basis von Intel Xeon „Ice Lake“-CPUs wird etwa die zehnfache Spitzenleistung des aktuellen Arbeitspferdes „Meggie“ bieten.

Inter-institutional cooperation to develop reusable APIs may be hard, especially for niche use-cases; but it would ease collaboration (when joining or replacing components), as well as healthy competition (less effort to compare implementations or machines).

This article described work financed by PRACE-6IP, under Grant agreement ID: 823767, under Project name LyNcs. The author wishes to thank colleagues or project partners who read the draft and gave advice.

Michele Martone

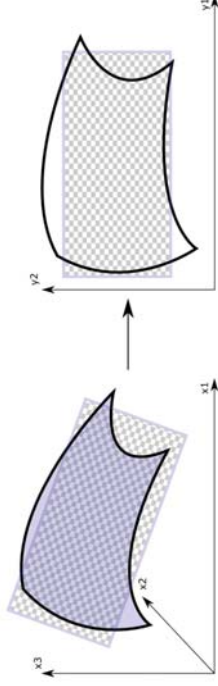


Figure 1: Manifold learning for dimensionality reduction produces a new set of coordinates (y_1, y_2) where the data originally stored with (x_1, x_2, x_3) coordinates is embedded.

datafold, this is possible through numerical approximation of eigenfunctions of the Laplace-Beltrami operator from typically very high-dimensional data (up to thousands of dimensions for image data) with unstructured point distribution. The Laplace-Beltrami operator is a core object of study in harmonic analysis. It is intimately connected to stochastic calculus, its eigenfunctions have a multiscale structure that is related to spatial discretization schemes such as sparse grids, it is the main object employed by spectral approximation and classification methods, and the generator of Gaussian (diffusion) processes. For manifold learning, the operator is often approximated by a kernel matrix, where its eigenfunctions are available as point evaluations in the eigenvectors of the matrix. The eigenfunctions can then be used as coordinates for a low-dimensional embedding or as a truncated basis of a function space (see Figure 2).

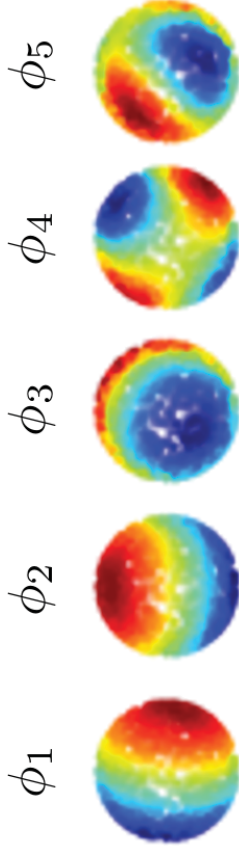


Figure 2: Five eigenfunctions of the Laplace-Beltrami operator on the sphere, approximated with Diffusion Maps and datafold. For the sphere, the eigenfunctions are known as “spherical harmonics”.

One of the most recent algorithms to approximate the operator and its eigenfunctions on scattered, high-dimensional data sets is called “Diffusion Maps” [3] from Coifman and Lafon. Other examples are the general “Kernel Principal Component Analysis” from Bengio et al., “Local Linear Embedding” from Belkin and Niyogi, or “Hessian Eigenmaps” from Donoho and Grime. A variety of manifold learning models already exist in the scikit-learn Python package. In addition to these, datafold provides an efficient implementation of “Diffusion Maps”. The implementation includes an optional sparse kernel matrix representation that allows it to scale to much larger datasets compared to scikit-learn. Datafold also supplies functionality for follow-up aspects of non-linear manifold learning, such as estimating the kernel scale parameters to describe the locality of points in a dataset, handling time series and extending the embedding to unseen data. This important map from the data to the new embedding space is referred to as “out-of-sample” extension. The eigenvectors, containing function values on the point cloud of the training data, are then extended to the point’s neighborhood regions. Out-of-sample extensions are, therefore, required to handle large input dimensions and interpolating general function values on point clouds. In datafold, out-of-sample extensions are implemented efficiently so that interpolated function values for millions of points can be computed in seconds on a standard desktop computer.

The Sparse BLAS API doesn’t cover all usage needs, but most of the recurring ones: this is why LIBRSB is dual-API. LIBRSB has been conceived to support Sparse BLAS from the beginning; the necessary groundwork eased writing accessors for FORTRAN, C++, and for interpreted languages PYTHON⁷ and GNU OCTAVE⁸.

Recent work in LIBRSB⁹ addressed re-engineering the *kernels* for multiplying a matrix by multiple vectors *at once* (BLAS_*usmm), so to improve the memory access and thus, performance. Such operation’s optimal context is that of *block Krylov methods*¹⁰. It’s important to mention that the API has **not** been changed while improving this functionality: only the internals did. A key aspect of LIBRSB is its polyalgorithmic nature: the library chooses a specific matrix representation (and algorithms) based on the given *sarsity pattern*. That allows to cover matrices from several domains (though focus has mainly been on *large* problems). It is likely that further *corner cases* (matrix domain, hardware) can only be efficiently tackled by expanding the polyalgorithmic character, perhaps combining with different libraries.

Computing architectures and APIs are diverging¹¹; APIs become obsolete; HPC is no exception here (Sparse BLAS itself needs a refresh). A consequence of this is a multiplicity of projects or research lines with scopes close or complementary to LIBRSB’s. The most notable such effort based in Germany is GINKGO, by Anzt et al.¹². While one may observe that different projects prioritize different goals (and it’s also good so!), probably a higher degree of reuse and interoperability would be beneficial. The community of library users has a role here: **requesting** library authors (in some cases, these are project partners) and vendors to follow standardized APIs is a first step to get good quality community software (and proprietary alike) being developed.

⁷PyRSB: <https://doi.org/10.25080/majora-1b6fd038-00e>

⁸sparsersb: <https://gnu-octave.github.io/packages/sparsersb>

⁹E.g. see <https://doi.org/10.5281/zenodo.5223873>

¹⁰See Agullo, Giraud, Jing in <https://doi.org/10.1137/140961912>.

¹¹See Thompson et al’s review in <https://doi.org/10.1145/3430936>

¹²GINKGO home page: <https://ginkgo-project.github.io/>

If an implementation is licensed as FLOSS⁴ (or even better, *free*), users can even enjoy the *right to repair*⁵ it. Availability of now easier to develop compliance-testing software shall in turn decimate bugs, improving software quality.

A section of the BLAS API specifies routines for *sparse algebra* mostly useful in the context of *iterative methods*. *Sparse matrices* often arise where problem topologies are irregular. In recent decades, computers' performance is increasingly negatively affected by irregular data accesses. This has stimulated efforts to develop techniques to overcome that. The LIBRSB is a project⁶ spun-off from a PhD research in that field, targeting cache-based CPUs (like your laptop, or your supercomputer's CPU). LIBRSB takes its name from the RSB (*Recursive Sparse Blocks*) matrix layout it uses. RSB exploits *cache locality* and *coarse-grained parallelism*; for best results, the memory layout can be *autotuned* for a specific operation.

LIBRSB is usable via different APIs: a native one (`rsb.h`), and SPARSE BLAS (in `C`, `blas_sparse.h`). The function naming scheme is similar to that of its *dense* counterpart: `'BLAS_Type'us'Op`, where *Type* is one of `d`, `s`, `z`, `c`, and `'us'` stays for *unstructured sparse*. Differently than in dense BLAS, function signatures do not imply any specific data structure: this is left to the implementation. A matrix is identified via an *integer-typed handle*. For instance, one can begin construction of a *double-precision* matrix with `BLAS_duscr_begin`, then populate and assemble it (notice the stateful technique), and only then use it in linear algebra operations like `BLAS_dusmv ('mv'` is short for *matrix-vector multiplication*). All this separates effectively interface and implementation, while not excluding further control over the matrix: function `BLAS_usp` allows setting *matrix properties*, which may be easily extended to support further implementation-specific functionality (LIBRSB does that).

⁴Free/Libre/Open Source Software – <https://www.gnu.org/philosophy/floss-and-foss.en.html>

⁵<https://www.fsf.org/blogs/community/watch-fight-to-repair-demand-the-right-to-repair>

⁶The LIBRSB home page: <http://librsb.sf.net>

Time-Series Analysis with the Koopman operator

The work of Koopman and von Neumann on ergodic, mixing, and chaotic dynamical systems revealed that there is a canonical, linear operator associated to each system [4,5]. This “Koopman operator”, albeit linear, captures the dynamics of nonlinear systems, and allows researchers to predict the evolution of functions defined on the system’s state space, referred to as “observables”. Linearity of the operator makes it amenable to finite-dimensional matrix approximations on computers, which has been heavily exploited in the last 20 years [6]. Particularly interesting complex-valued observables are the eigenfunctions of the Koopman operator. Together with the associate eigenvalues and modes, these allow researchers to gain insight to the modeled system. Essentially, the Koopman operator is to dynamical systems what the Laplace operator is to geometry: these operators encode everything about the system or object under study, and numerical approximations of them can therefore be used to access, process, store, and analyze all properties that are important to the respective application and, therefore, enable “scientific machine learning”. For linear dynamical systems, the Dynamic Mode Decomposition (DMD) algorithm efficiently approximates the Koopman operator from times series data. One extension of DMD to nonlinear systems with pure-point spectrum is Extended DMD (E-DMD). E-DMD approximates the Koopman operator with a matrix, based on a finite set of functions evaluated on the available data, the so-called “dictionary”. Many more numerical methods exist, such as Generalized Laplace Averaging, E-DMD with dictionary learning through neural networks, and dictionary selection through L1-optimization. The E-DMD framework can also be coupled with neural networks. One way is to learn the dictionary of E-DMD, while the other way is to integrate E-DMD in the training of a neural network to improve its convergence rate. In the latter case, the iterative training process itself can be interpreted as a dynamical system [7]. Datafold provides several implementations of the algorithms DMD and E-DMD. Finding a good choice for the dictionary in E-DMD is comparable

to the machine learning task of “model selection” and requires great flexibility in setting up the data processing pipeline. The flexibility of setting an arbitrary dictionary combined with a selection of the provided DMD variants is a core feature of datafold’s implementation. For example, if the state space of the dynamical system is a manifold, the processing pipeline enables its users to define manifold learning, dimensionality reduction, dictionary definition, and Koopman operator approximation as subsequent steps. Then, the entire pipeline can be fitted to the original data with one single command.

Datafold software

Datafold is an open-source software with a design that reflects a workflow hierarchy: from low-level data structures and algorithms to high-level meta-models intended to solve complex machine learning tasks. The key benefit of datafold is that it accommodates and integrates models on the different workflow levels. Each model has been investigated and tested individually, and many of datafold’s model implementations are already being used by research groups in the scientific community. In datafold, these models can be used in a single processing pipeline. Our integrated workflow facilitates the application of data-driven analysis and thus has the potential to boost widespread utilization. The implemented models are integrated into a software architecture with a clear modularization and an API that is templated from the scikit-learn project, which can be used as part of its processing pipeline. The data structures are subclasses from common objects of the Python scientific computing stack, allowing models to generalize for both static point clouds and temporally ordered time series collection data. The software design and modularity in datafold reflects two requirements: high flexibility to test model configurations, and openness to new model implementations with clear and well-defined scope.

We want to support active research in data-driven analysis with manifold context and thus target students, researchers and experienced practitioners from different fields of dataset analysis. The number of students working on additional features of datafold as their bachelor’s and master’s theses is

librsb: Supporting Shared-Memory Parallel Sparse BLAS from C/C++ up to Interpreted Languages



Leibniz-Rechenzentrum
der Bayerischen Akademie der Wissenschaften

“The dissemination of knowledge is of obvious value — the massive dissemination of error-loaded software is frightening.”

Edsger W. Dijkstra at the NATO Software Engineering Conference, 1968

When recompiling our computational code on the latest CPU-based machine (be it our laptop or a Tier-0 machine), we all take for granted that at least one high performance linear algebra implementation will be there for us, and accessible via the BLAS (Basic Linear Algebra Subroutines) API (Application Programming Interface). The bulk of the BLAS API has been discussed and agreed upon by an international committee of interested people over the course of the 1990s, during *BLAS Technical Forum* (BLAST)¹ meetings. That effort bears fruits we still enjoy today. Work is under way to extend specifications to further scenarios (e.g. GPUs, the C++ standard²).

The main advantage of a standardized API is that it makes code portability across platforms trivial; but also within the same platform, it allows several implementations to compete. The benefits are manifold: for a team developing a scientific code, it spares the efforts of learning different APIs for the same task³; for a computing center, helps avoiding *customer lock-in*. Even the implementors’ job is better-posed, when aiming at comparability with a competitor’s product.

¹BLAS Technical Forum: <http://www.netlib.org/blas/blast-forum/>

²See Hoemmen et al. in <http://www.open-std.org/jtc1/sc22/wg21/docs/papers/2019/p1417r0.pdf> and <http://www.open-std.org/jtc1/sc22/wg21/docs/papers/2021/p1673r4.html>

³Another example is the work which went into developing the OPENMP standard over the last 23 years. That experience has recently inspired work of OPENMP extensions for PYTHON – a promising idea.

Mit den Wirtschaftswissenschaften ging es vor wenigen Jahren los. Von den dreizehn vorgesehenen Professuren sind die meisten inzwischen besetzt, die entsprechenden Studiengänge laufen und erfreuen sich gerade auch international rapide wachsender Beliebtheit. Und nun läuft also „Heilbronn 2.0“ an – der TUM Campus Heilbronn bekommt nach der ökonomischen nun auch eine technologische, eine digitale Komponente. Ende August wurden neun weitere Professuren ausgeschrieben (in Kurzform 3x theoretische, 3x technische und 3x praktische Informatik), ein Bachelor und ein Master „Information Engineering“ wurden eingerichtet, und der Bachelor läuft jetzt im Wintersemester als Pilotversion (zunächst über Vertretungsprofessuren realisiert) an. Trotz der extrem kurzen Vorlaufzeiten übrighens mit Bewerbungszahlen, die schon für den Piloten eine dreistellige Kohorte erwarten lassen.

Die Bewältigung der neun Berufungsverfahren neben dem aktuell ohnehin stattfindenden Informatik-Ausbau ist natürlich eine Herausforderung. Wir organisieren das in drei Berufungsausschüssen, die jeweils drei der Verfahren gemeinsam abwickeln, jeweils unter dem Vorsitz eines Zweitmitglieds unserer Fakultät. Neben anderen Vorteilen ermöglicht dieses Vorgehen eine optimale Abstimmung zwischen den einzelnen Verfahren und somit eine stringente Besetzung der Professuren. Soweit der Plan – die nächsten Monate werden zeigen, wie schnell und wie gut er aufgeht.

Parallel dazu wird an einer tragfähigen Governance gearbeitet – damit aus zwei hintereinander auf den Weg gebrachten Großprojekten wirklich ein TUM Campus Heilbronn wird. A propos zwei – natürlich steht das Grobgerüst von „Heilbronn 3.0“ bereits, hier sitzen aktuell die involvierten Anwaltskanzleien über den Vertragsentwürfen. Schon spannend, was so alles passiert. Ob das Quartl aufgrund dieser Aktivitäten im Schwäbischen zum „Quartele“ mutiert oder zumindest einen einschlägigen Ableger bekommt? Nun, nichts erscheint als unmöglich in diesem Zusammenhang.

Hans-Joachim Bungartz.

already quite large. This is in part because of the media hype surrounding AI, but also shows that students are interested to understand the background and theory behind the algorithms involved.

If you are interested in datafold, either to use it or to extend its functionality, please get in touch!

Daniel Lehmborg
Felix Dietrich

References

1. Datafold website: <https://datafold-dev.gitlab.io/datafold/index.html>
2. Lehmberg et al., (2020). datafold: data-driven models for point clouds and time series on manifolds. *Journal of Open Source Software*, 5(51), 2283, <https://doi.org/10.21105/joss.02283>.
3. Coifman, R. R. & Lafon, S. Diffusion maps. *Applied and Computational Harmonic Analysis*, Elsevier BV, 2006, 21, 5-30.
4. Koopman, B. O. Hamiltonian systems and transformation in Hilbert space. *Proceedings of the National Academy of Sciences of the United States of America*, 1931, 17, 315-318.
5. Koopman, B. O. & v. Neumann, J. Dynamical Systems of Continuous Spectra. *Proceedings of the National Academy of Sciences*, *Proceedings of the National Academy of Sciences*, 1932, 18, 255-263.
6. Mezić, I. Spectral Properties of Dynamical Systems, Model Reduction and Decompositions. *Nonlinear Dynamics*, 2005, 41, 309-325.
7. Dietrich, F.; Thiem, T. N. & Kevrekidis, I. G. On the Koopman Operator of Algorithms. *SIAM Journal on Applied Dynamical Systems*, *Society for Industrial & Applied Mathematics (SIAM)*, 2020, 19, 860-885.

IN.TUM goes Heilbronn



Nachdem Ende Juli – ziemlich genau ein Jahr nach der Unterzeichnung der Erweiterung des Kooperationsvertrags zwischen der Dieter-Schwarz-Stiftung und der TU München – auch die letzten juristischen Bedingungen erfüllt waren, geht es jetzt richtig los mit den Aktivitäten der TUM-Informatik auf dem Bildungscampus Heilbronn.

Dass ein Engagement einer deutschen Universität in einem anderen Bundesland wesentlich ungewöhnlicher als ein ebensolches in Qatar, Singapur oder in Afrika ist, mag den einen oder die andere verwundern, liegt aber natürlich in dem eisernen deutschen Grundsatz „Bildung ist Länderspezifität“ begründet. Gut, so ganz eisern ist das alles auch nicht mehr seit den letzten diesbezüglichen Grundgesetzänderungen, aber das Prinzip steht. Insofern war und ist auch das Engagement der TUM auf dem Heilbronner Bildungscampus etwas Besonderes und wissenschaftspolitisch durchaus nicht un-delikat. Ermöglicht wird der Campus vor allem durch die Vision und das Engagement der Dieter-Schwarz-Stiftung, die über eine langfristig, breit und tief angelegte Förderung der Region einen nachhaltigen Impuls geben möchte.

Schauen wir uns die drei Adjektive an, die allesamt Bemerkenswertes beschreiben. Erstens langfristig – kein Anschub, kein „ich geb dir mal für drei Jahre was, und danach schau selbst, wie es weitergeht“, sondern eine dauerhaft (bzw. durch rechtliche Rahmenbedingungen zunächst auf 30 Jahre) angelegte Zusammenarbeit. Zweitens breit – alle Phasen der Bildung, alle Hochschultypen ebenso wie außeruniversitäre Forschung, Lokales, Nationales und Internationales sind integriert. Drittens tief – es geht nicht um Auftragsforschung oder ein „Studienprogramm light“, sondern um das volle Portfolio, das eine Uni wie die TUM betreibt, nur eben in Heilbronn.