WWU MÜNSTER

Nuclear norms in the context of dynamic MRI:
Regularization techniques and asymptotic analysis

Doctoral Thesis

Meike Kinzel

- 2021 -

living.knowledge

MM Mathematics Münster

Nuclear norms in the context of dynamic MRI:

Regularization techniques and asymptotic analysis

Inaugural Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften

– Dr. rer. nat. –

im Fachbereich Mathematik und Informatik
der Mathematisch-Naturwissenschaftlichen Fakultät
der Westfälischen Wilhelms-Universität Münster

Fach: Mathematik

eingereicht von

Meike Kinzel

– 2021 –

Dekan:	Prof. Dr. Xiaoyi Jiang
Erster Gutachter:	Prof. Dr. Martin Burger Friedrich-Alexander-Universität Erlangen
Zweiter Gutachter:	Prof. Dr. Benedikt Wirth Westfälische Wilhelms-Universität Münster
Tag der mündlichen Prüfung:	10.11.2021
Tag der Promotion:	10.11.2021

Abstract

This thesis sheds new light on the reconstruction of dynamic MR data via variational methods that regularize with nuclear norms.
On the one hand, we devote ourselves to the classical nuclear norm and discuss its benefits in connection with the reconstruction of dynamic series that are presumed to consist of highly correlated frames. In this context we also present how its variational implementation, in combination with the ℓ^1-norm, enables to automatically track cells in MR scans. On the other hand, we consider the nuclear norm in a broader sense. Concentrating on linear mappings between non-euclidean vector spaces, we derive a generalized version. Due to its adaptivity, this opens up new application-oriented possibilities. Focusing on the tasks in dynamic MRI, we use the more general framework to deduce an approach which incorporates a-priori knowledge on the occurence of smooth dynamics into the process of reconstruction. In a second part we then contemplate the above mentioned approaches from a theoretical point of view. Aiming for continuous variational problems that mimic the infinitely fine temporal and/or spatial resolution of an MR scanner and represent the discrete ones in an appropriate manner, we address the study of their Γ-convergence. In doing so, we show that the considered discrete nuclear norms Γ-converge toward their natural continuous counterparts.

Acknowledgements

Before turning to the serious part of this thesis, I would like to take the chance to thank all the people without whom this work would not have been possible:

First of all, I want to express my gratitude toward all the colleagues and members of the department that made my PhD the exciting and hilarious journey it was. Being part of the infamous ‘Dönerbude’ and having so many great companions has seriously meant a lot to me. Thank you,

- *Martin Burger*, for introducing me to the fascinating field of mathematical image processing, for putting your trust in me and giving me so much freedom, for never getting impatient when it, once again, took some time until I got your point, for enabling me to travel the world visiting several international conferences, for having a knack for employing nice people, for teaching me that mails containing three letters are a valid tool of communication, and for being the coolest MC a math department has ever seen.
- *Daniel Tenbrinck*, for operating against my self-doubts and encouraging me to start my PhD, for your commitment regarding the NoMADS project and the associated chance to spend a summer in Italy, for always keeping the connection between Münster and Erlangen - even if this was requiring the introduction to ‘car bombs’, and for a Sexy-and-I-know-it performance I will never be able to forget.
- *Leonard Kreutz*, for being my secret unofficial co-advisor, for your great(!!!) mathematical advice, for proofreading a part of this thesis, for making me feel less stupid while presenting your very special applicational skills, for sharing my nerdiness about the Italian language, for introducing me to Chillhop, but most of all for being so untiringly patient while enduring all of my rides on the mood swing. I am not sure whether I would have ever submitted this work without you.
- *Julian Rasch* (aka DJ Ulian), for sympathetically sharing all the ups and downs I faced on my way to this thesis, for your effort to bail me out, for proofreading a part of this work, for revealing ‘the trick’ to me, for being my very personal career service and giving me so much good advice, for providing me with a lot of goodies from the Raschian library, for always serving the best and most professionally looking coffee, and last but not least for being so damn smart and still not taking this whole thing too serious.
- *Fjedor Gaede*, for introducing me to the working group during my master thesis, for those good old sweatpants days in our Dönerbuden office when you were dropping ‘Meike mach’ du mal auf, ich hab’ grad keine Hose an’, for tons of songs about my name, for your catching laugh, for teaching me how to play that emoticon game strong (`U+1F529`), for

your smashing break dance performance, for proofreading a part of this thesis, but most importantly for being a really good friend, who always knew when I needed a hug.

- *Ina Humpert*, for your flawless Italian skills, for letting lush greenery rule our office, for being the creative mind behind all our exaggerated gifts, for bringing back that Sims vibe and keeping track of our needs, fears and wishes, for always making sure to feed my sweet tooth, and for being so incredibly considerate during our last weeks at the department.

- *Ramona Sasse*, for being so genuinely you, for all your effort encouraging kids, women or anybody else to engage with the fascinating field of applied mathematics, for always having a sympathetic ear for any kind of rant, and for keeping me awake being the French press guru numero uno (I mean seriously, did I ever prepare a coffee by myself? – sorry for that!).

- *Lea Föcke*, for implementing that genius function which saved my life (`jfscrollimage` for president!), for always being there to help out with any kind of technical issues, for hosting the coolest parties I have ever visited, and for a chipmunk version of 'Uptown funk' I still get nightmares about.

- *Ulrich Böttcher*, for providing me with this template, for being the most reliable resource when it comes to weather forecasts (or basically anything else), for a good amount of sweaty hours together at the gym, for some more or less successful evenings fighting against Gors, Skrals and dragons just in order to save that drunkard Prince Thorald, and for convincing us that the computer games museum in Berlin is a must-see.

- *Carolin Dirks*, for your untiring positive energy, for that great Wizard-Party we hosted while being stuck in Hannover, and for all of the freaking delicious treats you were serving at our cake Mondays.

- *Eva-Maria Brinkmann*, for taking me to my very first symposium and introducing me to the imaging community, for being the most considerate and kind person I have ever met, and for being the mother the Dönerbude has needed.

- *Frank Wübbeling*, for organizing the world's most fabulous skiing seminar, for being our very personal 'Gossip Girl', and for never forgetting to take care of us 'Leftovers'.

- *Claudia Giesbert*, for taking all that organizational stuff off my shoulders, while being the most affectionate secretary one could ask for.

- *Lorenzo Rosasco*, for hosting me at the University of Genoa and making me feel like a real member of your working group, for teaching me a lot about machine learning, and for always entertaining with the world's biggest collection of statement shirts.

- *Luigi Carratino* and *Fabio Anselmi*, for all your effort introducing that German girl to 'la dolce vita', for tons of 'splashy' and 'shakeratory' coffee breaks, and in general for making Frigolab feel so much warmer than it actually was.

Likewise, I want to thank all of my family and friends who supported me in a non-professional way. Your encouraging words and distracting efforts always pushed me further and kept me sane. However, my special gratitude goes to

- *Marvin Sandkuhle*, for patiently consoling me everytime I felt overwhelmed with my to-do's, for buying that insane enormous screen that I insisted on not needing but loved in the end, for always making sure to get me 'out of that matrix', and for making me laugh. every. single. day. There is seriously no company that I would prefer while being locked up in a pandemic.
- *Steffen*, *Andrea* and *Wolfgang Kinzel*, for your unconditional support in everything I do. I am so damn proud to call you my family, you rock!

This work was supported by the German Federal Ministry of Education and Research under the project ID 05M16PMB (MED4D) and the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 777826 (NoMADS).

Contents

1 Introduction **1**

2 Mathematical preliminaries **7**
2.1 Function spaces 7
2.2 Integral operators 10
2.2.1 Singular value decomposition 11
2.3 Γ-convergence 15

3 Inverse problems and variational methods **19**
3.1 Inverse problems 19
3.2 Variational methods 22
3.2.1 Bayesian modeling 23
3.2.2 Transition to the infinite-dimensional case 26
3.2.3 Example: Sparse signal reconstruction 26
3.2.4 Existence of solutions 28

4 Variational methods for dynamic MRI **33**
4.1 Dynamic MRI: Physics and methodical limits 33
4.1.1 The measurement process 33
4.1.2 The (ill-posed) inverse problem to dynamic MRI 35
4.2 Matrix norms in the context of dynamic MRI 38
4.2.1 Mixed norms 38
4.2.2 Nuclear norm 39
4.2.3 Generalized nuclear norm 41
4.3 Variational methods for the reconstruction of dynamics 44
4.3.1 Linear dependence among frames 44
4.3.2 Superposition of background and dynamics 47
4.3.3 Linear dependence favoring smooth dynamics 49

5 Numerical implementation and results **55**
5.1 Algorithmic solution 55
5.2 Computational experiments 63
5.2.1 Undersampling imposing a superposition of dynamics 63
5.2.2 Tracking of dynamics 65
5.2.3 Undersampling imposing the occurrence of few smooth dynamics? 67

6 Asymptotic behavior **71**
6.1 Analytical study of general energy functionals 72
6.1.1 Formal setting 72

6.1.2 Coercivity 78
6.1.3 Continuity of data fidelity terms 81
6.2 Γ-convergence of regularizing norms 82
6.2.1 Mixed norms 83
6.2.2 Nuclear norm 87
6.2.3 Generalized nuclear norm 104

7 Conclusion and outlook **129**

Bibliography **133**

List of figures and tables **139**

1 Introduction

In 1952 F. Bloch and E. M. Purcell received the Nobel Prize for ‘their development of new methods for nuclear magnetic precision measurements and discoveries in connection therewith’. Their work allowed P. Lauterbur to develop the technique of *magnetic resonance imaging* (MRI) [Lauterbur, 1973], which revolutionized the way medical diagnosis are determined today. MRI is an imaging method to display structure and functionality of soft tissue and organs in the human body. In contrast to hardened body parts, such as bones, these contain hydrogen nuclei which, due to their spin, generate an electric field. Exploiting this fact, in MRI a combination of static and high-frequency magnetic fields allows to excite these atoms. By virtue of their excitation, the nuclei move in different ways, depending on the local tissue they belong to. This movement induces an electrical current, and thus allows to acquire measurements from which it is possible to extrapolate an image depicting the physiological origin. Beyond the generation of a single static image, this procedure can also be repeated several times to obtain a series of outcomes. This is called *dynamic MRI* and serves to image dynamic processes. In the context of *functional MRI*, it can for example be applied to study neuronal activity and therefore allows to characterize and detect brain diseases (see, e.g., [Huettel et al., 2004]). In *perfusion MRI* it is used to visually represent and quantify the blood flow in organs (see, e.g., [Petrella and Provenzale, 2000]). Beyond that, the technique is applied in scenarios where one is interested in tracking the movement of cells in the human body (see, e.g., [Hemmer et al., 2015; Sánchez et al., 2012; Masthoff et al., 2019]).

Each of these procedures provides a non-invasive possibility to get full access to anatomical information. However, all of them feature one major drawback: the process of acquiring raw data is very time consuming. This is particularly obstructive in the context of dynamic MRI, since its success strongly depends on the ability to perform measurements in short time frames. A natural resort to circumvent this obstacle is to perform fewer measurements. In order to understand how this affects the quality of the involved images, it is instructive to look at the reconstruction from a mathematical point of view.

Within the modeling assumptions of MRI, it is presumed that the measured data is generated by a physiological origin. Therefore, in an idealized setting, the process of data acquisition is commonly expressed by an equation of the form

$$Ky = x. \tag{1.1}$$

Here, y describes the intensity of the grayscale image, which represents the object under consideration, while x characterizes the measured data. The involved operator K serves to model the transition from one to the other and, in the present setting, is given as the *fourier transform* (see, e.g., [Elster and Burdette, 2001]). As this operator is bijective, equation (1.1) in general allows to recover the desired image y from data x by applying the inverse of K.
However, as described above, it would be desirable to only perform a reduced amount of measurements. Hence, if recovering y is still to lead to an image of high resolution, the included operator needs to be modified in order to account for the absent data. This leads to a version that does not posses a continuous inverse anymore. Combined with the fact that, in practice, measurement devices do not work perfectly but generate noisy data, this entails a severe problem: the recovery of y constitutes an *ill-posed inverse problem*. This means that inevitable small measurement errors may lead to large errors in the reconstructed image.
The study of such ill-posed inverse problems is subject to vibrant research, both from an applied and theoretical point of view (see, e.g., [Groetsch, 1993; Colton et al., 2012; Tikhonov et al., 1987]). With regards to MRI, a point of emphasize is to compensate the absence of sufficient data by involving a-priori information on the expected reconstruction. To do so, one considers a modified problem by either restricting the admissible set of solutions or changing the reconstruction procedure to favor the anticipated outcome. One way to explicitly implement this modification is to follow the framework of *variational modeling*. This relies on the observation that (under suitable conditions) finding a solution to (1.1) is equivalent to determining the minimizer of

$$y \mapsto \frac{1}{2}\|Ky - x\|^2. \tag{1.2}$$

In comparison to the algebraic approach, the variational one presents two advantages. On the one hand, it guarantees the existence of solutions. Thus, even if, as a result of measurement errors, (1.1) does not admit an exact solution, in most cases an approximate one can be found via (1.2). On the other hand, one can easily manipulate the desired solution, by including an additional term, that encodes the a-priori knowledge. Mathematically speaking, that means to consider a modified minimum problem of the form

$$y \mapsto \frac{1}{2}\|Ky - x\|^2 + \alpha\mathcal{R}(y). \tag{1.3}$$

Here, the additional term $\mathcal{R}$ is called *regularizer* and is used to penalize undesirable properties of possible solutions. The parameter $\alpha > 0$ serves to balance the importance of the two terms involved. In this setting, appropriate choices of $\mathcal{R}$ entail the benefit that the reconstruction y depends continuously on the data x.
In dynamic MRI, respecting that the recorded measurements for different time steps are all based on the same anatomic structure, one expects that the reconstructed frames of the time series are highly correlated. Recalling our previous discussion about (1.3), it therefore stands to reason to include this observation as a-priori knowledge in the reconstruction process. However,

in order to realize this idea, a suitable regularizer $\mathcal{R}$, favoring this strong linear dependence in time, has to be defined. Taking into account the expected outcome, one can assert that the matrix, storing the frames in a vectorized version, has low rank. That means that only few of its singular values differ from zero, or in other words, it has a sparse *singular value decomposition* (SVD). In general, to enforce sparsity of solutions for optimization problems, penalizing the ℓ^1-norm turned out to be a useful tool (see [Tibshirani, 1996]). This stems from the fact that it is the convex relaxation of the ℓ^0-norm (see [Donoho and Elad, 2003]), which measures the cardinality of non-zero entries and therefore favors sparsity as far as minimization is concerned. In a similar fashion, aiming for few non-zero singular values, penalizing their ℓ^1-norm is a canonical approach. Using this so-called *nuclear norm* as a regularizing term has been subject to investigation in connection with various applications (see [Candès and Recht, 2009; Candès and Tao, 2005; Recht et al., 2010]). In the framework of dynamic MRI, this idea has been implemented, for example, with applications to contrast enhanced breast imaging (see [Liang, 2007; Haldar and Liang, 2010]). Beyond that, in [Lingala et al., 2011] and [Zhao et al., 2012] an adapted approach has been suggested. Here, referring to applications in cardiac perfusion MRI, the solution is not only expected to be of low rank, but also to have a sparse representation with respect to a suitable basis. In order to encode those two conditions at the same time, the regularizer $\mathcal{R}$ in (1.3) is chosen as a combination of the nuclear norm and an ℓ^1-norm in the respective basis. A different take on tackling dynamic MRI has been proposed in [Otazo et al., 2015; Gao et al., 2012]. In these works it is assumed that the desired reconstruction is the superposition of a low-rank matrix L, modeling the temporally correlated background, and a sparse component S representing the dynamics of interest. Therefore, the unknown in (1.3) is postulated to be the sum of L and S, while $\mathcal{R}$ penalizes the nuclear norm of L as well as the ℓ^1-norm of S. Here, the validity of the procedure has been demonstrated in connection with various experiments in MRI including perfusion, time-resolved angiography, and cardiac cine.

Motivated by this rich family of applications, the scope of this thesis is to investigate more thoroughly the nuclear norm as a regularizer, where we mainly put our focus on its generalization and the study of its asymptotic analysis.
More precisely, we revisit the aforementioned approaches with regard to the task of tracking cells in the human body. Since here, the object under consideration separates into an almost constant background and the small scale dynamics of interest, this assignment can be perfectly placed in the above class of problems. By performing numerical experiments, we show that this procedure proves successful for image reconstruction as well as automated cell tracking in previously reconstructed MR data.
Moreover, based on these results, we study the nuclear norm as a regularization term in a broader sense. For this purpose, we first of all observe that the SVD of a matrix, which encodes the time series of images, encapsulates its dynamics and their intensities. Therefore, as in the case of $K = \mathrm{Id}$ (i.e., in the setting of denoising) regularizing with the nuclear norm boils down to the linear shrinkage of the singular values of A (see [Cai et al., 2010]), minimizing (1.3) selects

the dominant dynamics. With that in mind, we consider modified SVDs - and thus modified nuclear norms - to manipulate this selection procedure. To this end, we note that the classical SVD of a matrix $A \in \mathbb{C}^{m\times n}$ is based on its interpretation as a linear mapping from $(\mathbb{C}^n, \langle\cdot,\cdot\rangle)$ to $(\mathbb{C}^m, \langle\cdot,\cdot\rangle)$, where $\langle\cdot,\cdot\rangle$ denotes the euclidean scalar product. Hence, to construct a modified SVD, we endow $\mathbb{C}^m$ and $\mathbb{C}^n$ with the inner products

$$\langle x_1, x_2\rangle_C := x_1^H C x_2,\ \forall x_1, x_2 \in \mathbb{C}^m \quad \text{and} \quad \langle y_1, y_2\rangle_D := y_1^H D y_2,\ \forall y_1, y_2 \in \mathbb{C}^n,$$

where $C \in \mathbb{C}^{m\times m}$ and $D \in \mathbb{C}^{n\times n}$ are Hermitian and positive definite matrices. The resulting singular values $(\sigma_\xi^{C,D})_\xi \subset \mathbb{R}$ are now obtained as the eigenvalues of the matrix $A^{*_{C,D}}A$, where $A^{*_{C,D}} \in \mathbb{C}^{n\times m}$ is the adjoint of A respecting the scalar products $\langle\cdot,\cdot\rangle_C$ and $\langle\cdot,\cdot\rangle_D$. Of course, since also the corresponding singular vectors are affected by the choice of the matrices C and D, in the light of the aforementioned, the captured dynamics as well as their magnitudes are changed. For appropriate choices of C and D, we use the resulting nuclear norm as a penalization term for (1.3). Contrary to the classical setting, we observe that, in the denoising scenario, this implementation leads to nonlinear shrinkage of the modified singular values. This also allows us to draw conclusions about the general problem in (1.3), where $\mathcal{R}$ is chosen as the classical nuclear norm. In particular, the nonlinear effects described above render the reconstruction unclear for the general problem and suggest that further investigation is needed.

Subsequently, we address the asymptotic behavior of the previously discussed minimization problems for very fine spatial and/or temporal resolution. This is relevant, because, even when reverting to efficient algorithms, the reconstruction procedure can be very time consuming as the degrees of freedom become large. By deriving ‘effective’ continuum problems, we tackle this question from an analytical point of view. Since these surrogate problems shall be good approximations of those in question, we resort to techniques of the *calculus of variations*. More precisely, we employ the concept of Γ-*convergence*. This proved to be a powerful tool to study the asymptotic behavior of sequences of minimization problems, since, together with *equi-coercivity*, it ensures the convergence of minimizers (see, e.g., [Braides, 2002]). Identifying the Γ-limit is a two step procedure: First, it has to be shown that the limit functional is an asymptotic lower bound; Secondly, one has to make sure that this bound is optimal.

Within this framework, in analogy with (1.3), we here examine

$$\mathcal{F}_\mu(A) = \frac{1}{2}\|KA - B_\mu\|^2 + \alpha\mathcal{R}_\mu(A), \tag{1.4}$$

where $\mu \in \mathbb{N}$ encodes the spatial and temporal dimensions of the matrix $A \in \mathbb{R}^{m_\mu \times n_\mu}$ and the regularizer $\mathcal{R}_\mu$ ranges between (weighted) mixed p, q-norms, with $p, q > 1$, the classical nuclear norm, and its generalized version. We show that this family of problems satisfies a suitable notion of equi-coercivity. More specifically, following a semi-discrete approach, we show that for sequences of matrices $(A_\mu)_\mu$ that satisfy $\sup_\mu\ F_\mu(A_\mu) < +\infty$, there exist suitable interpolations $(\hat{A}_\mu)_\mu$ and an integral operator A such that $\hat{A}_\mu$ converges to A with respect to the weak operator topology (see Chapter 6 for the precise choice of topology). With respect to this very topology, we then show that the Γ-limit of the mixed p, q-norm is given by the $L^{p,q}$-norm on the space of $L^{p,q}$-integral operators. Our result is the functional analytic analogue to [Heins, 2014, Thrm. 7.4] which is based on tools of measure theory. Beyond that, focusing on the classical nuclear norm, we prove that its Γ-limit is well defined on the space of $L^{2,2}$-integral operators and, for such operators, can again be represented as the sum of the singular values. Regarding the generalized nuclear norm, we consider the spaces $(\mathbb{R}^{m_\mu}, \langle \cdot, \cdot \rangle_{C_\mu})$ and $(\mathbb{R}^{n_\mu}, \langle \cdot, \cdot \rangle_{D_\mu})$. Here, $C_\mu \in \mathbb{R}^{m_\mu \times m_\mu}$ and $D_\mu \in \mathbb{R}^{n_\mu \times n_\mu}$ define scalar products, that are equi-continuous and equi-coercive with respect to the parameter μ. Moreover, we assume that the sequences $(C_\mu^{1/2})_\mu$ and $(D_\mu^{1/2})_\mu$ (now defined on L^2 via an embedding) strongly converge to limiting operators $C^{1/2}$ and $D^{1/2}$, respectively. Based on these hypotheses, we demonstrate that the generalized nuclear norm regarding C_μ and D_μ Γ-converges to the sum of generalized singular values characterized via C and D. To do so, we employ the following two key arguments: First, we relate orthonormal bases of $(L^2, \langle \cdot, \cdot \rangle_C)$, respectively $(L^2, \langle \cdot, \cdot \rangle_D)$, to the ones of $(\mathbb{R}^{m_\mu}, \langle \cdot, \cdot \rangle_{C_\mu})$, resp. $(\mathbb{R}^{n_\mu}, \langle \cdot, \cdot \rangle_{D_\mu})$, via the embedding and a Gram-Schmidt argument. Afterwards, we exploit the dual structure of the generalized nuclear norm to compare the sequence of functionals with the limit.

The thesis is organized as follows. In Chapter 2, we present the mathematical concepts that form the basis for the analytical part of this work. Chapter 3 gives a short introduction to the theory of inverse problems and discusses variational approaches to study them. Subsequently, in Chapter 4, we review how these approaches can be implemented in the context of dynamic MRI. To do so, we first detail the process of data acquisition. Secondly, we define different matrix norms and convey their relevance with respect to different applications. The numerical treatment of the resulting minimization problems is discussed in Chapter 5. Additionally, we employ the derived algorithms to perform computational experiments addressing the aforementioned applications. Chapter 6 is devoted to the asymptotic analysis of the mixed p, q-norm and the (generalized) nuclear norm. Finally, we conclude this work by pointing to potential future developments motivated by the problems considered in Chapters 4 – 6.

2
Mathematical preliminaries

Before diving into the main part of this thesis we want to start with the brief recall of some mathematical concepts that will play a relevant role within the subsequent considerations. While doing so, we especially focus on the aspects which will be important for the analytical part of this work.

2.1 Function spaces

In order to start, we first of all want to touch upon two kinds of classes of function spaces which basically can be interpreted as generalizations of the L^p-spaces. However, since these are usually encountered less frequently, we here want to give a proper definition and point to some useful properties. To do so we initially turn toward the introduction of the vector-valued counterpart of the classical L^p-spaces.

Definition 2.1. Let $\Sigma \subset \mathbb{R}^d$ be open, $n \in \mathbb{N}$ and $1 \leq p \leq \infty$. Then

$$L^p(\Sigma; \mathbb{R}^n) \coloneqq \{\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n) \,|\, \mathbf{x}_j \in L^p(\Sigma) \ \ \forall j \in \{1, \dots, n\}\}$$

together with

$$||\mathbf{x}||_{L^p(\Sigma;\mathbb{R}^n)} \coloneqq \left(\sum_{j=1}^{n} ||\mathbf{x}_j||^2_{L^p(\Sigma)} \right)^{\frac{1}{2}}$$

defines the normed vector space of all n-tuples of functions in $L^p(\Sigma)$.

Note that due to its composition of finitely many L^p-spaces we can benefit from a few nice properties and follow proof ideas that work well on the level of Lebesgue spaces. One of them leads to the identification of the dual space of $L^p(\Sigma; \mathbb{R}^n)$.

Remark 2.2. Let $1 < p < \infty$. Similar to the case of regular L^p-spaces, applying the inequalities of Hölder and Cauchy-Schwarz yields

$$\begin{aligned}\left|\sum_{j=1}^{n}\langle \mathbf{x}_j, \mathbf{x}'_j\rangle_{L^p(\Sigma),L^{p'}(\Sigma)}\right| &\leq \sum_{j=1}^{n} ||\mathbf{x}_j||_{L^p(\Sigma)}\,||\mathbf{x}'_j||_{L^{p'}(\Sigma)} \\ &\leq \left(\sum_{j=1}^{n}||\mathbf{x}_j||^2_{L^p(\Sigma)}\right)^{\frac{1}{2}}\left(\sum_{j'=1}^{n}||\mathbf{x}'_{j'}||^2_{L^{p'}(\Sigma)}\right)^{\frac{1}{2}} \\ &= ||\mathbf{x}||_{L^p(\Sigma;\mathbb{R}^n)}\,||\mathbf{x}'||_{L^{p'}(\Sigma;\mathbb{R}^n)}\end{aligned}$$

for $\mathbf{x} \in L^p(\Sigma;\mathbb{R}^n)$, $\mathbf{x}' \in L^{p'}(\Sigma;\mathbb{R}^n)$ and p' being the conjugate exponent to p. Thus $T_{\mathbf{x}'} : L^p(\Sigma;\mathbb{R}^n) \to \mathbb{R}$ characterized by

$$T_{\mathbf{x}'} : \mathbf{x} \mapsto \sum_{j=1}^{n}\langle \mathbf{x}_j, \mathbf{x}'_j\rangle_{L^p(\Sigma),L^{p'}(\Sigma;\mathbb{R}^n)}$$

is linear and bounded and therefore an element of the dual space to $L^p(\Sigma;\mathbb{R}^n)$ with

$$||T_{\mathbf{x}'}|| \leq ||\mathbf{x}'||_{L^{p'}(\Sigma)^n}. \tag{2.1}$$

Following the exact same reasoning as in [Dunford and Schwartz, 1988, Chap. IV.8, Thrm. 1] for the L^p-spaces then leads to the insight that every linear and bounded functional on $L^p(\Sigma;\mathbb{R}^n)$ is of this form and that the norms in (2.1) even coincide. Consequently $L^{p'}(\Sigma;\mathbb{R}^n)$ can be identified with the dual space of $L^p(\Sigma;\mathbb{R}^n)$ and the reflexivity of both is manifested.

Based on this reflexivity, also the following property on the weak convergence in $L^p(\Sigma;\mathbb{R}^n)$ can directly be inferred from its classical counterpart.

Corollary 2.3. The space $L^p(\Sigma;\mathbb{R}^n)$ is weakly sequentially complete.

In addition to that, when defining $C_c^\infty(\Sigma;\mathbb{R}^n)$ likewise, i.e., as the space of n-tuples of functions in $C_c^\infty(\Sigma)$, we can easily transfer the well-known density statement to $L^p(\Sigma;\mathbb{R}^n)$.

Corollary 2.4. Let $1 \leq p < \infty$. Then, $C_c^\infty(\Sigma;\mathbb{R}^n)$ lies dense in $L^p(\Sigma;\mathbb{R}^n)$.

With that, we want to turn toward an even more generalized version of the classical Lebesgue spaces introduced in the 1960s by Benedek and Panzone.

Definition 2.5. (cf. [Benedek and Panzone, 1961, Sec. 1])
Let $1 \leq p, q \leq \infty$ and $\Sigma \subset \mathbb{R}^d$, $\Omega \subset \mathbb{R}^{d'}$. Then, the *mixed Lebesgue space* $L^{p,q}(\Sigma \times \Omega)$ is

defined as the set of all Lebesgue-measurable functions $t : \Sigma \times \Omega \to \mathbb{R}$, with

$$\|t\|_{L^{p,q}(\Sigma\times\Omega)} := \left(\int_\Sigma \left(\int_\Omega |t(s,r)|^q \, \mathrm{d}r \right)^{\frac{p}{q}} \mathrm{d}s \right)^{\frac{1}{p}} < \infty.$$

It was found that also this generalization shares a lot of properties of the L^p-spaces. One of them concerns the corresponding dual space.

Theorem 2.6. (cf. [Benedek and Panzone, 1961, Sec. 4, Thrm. 1 & Lem. 1])
Let $1 < p, q < \infty$. Suppose that p' and q' denote the conjugate exponents to p and q. Then, $L^{p,q}(\Sigma \times \Omega)$ is reflexive and its dual space $(L^{p,q}(\Sigma \times \Omega))^*$ can be identified with $L^{p',q'}(\Sigma \times \Omega)$.

Beyond that, as was to be expected, also in this case the associated dual norm coincides with the norm on $L^{p,q}(\Sigma \times \Omega)$.

Theorem 2.7. (cf. [Benedek and Panzone, 1961, Sec. 2, Thrm. 1])
Let $1 \leq p, q \leq \infty$ and $t \in L^{p,q}(\Sigma \times \Omega)$. Then,

$$\|t\|_{L^{p,q}(\Sigma\times\Omega)} = \sup_{\|g\|_{L^{p',q'}(\Sigma\times\Omega)}=1} \int_{\Sigma\times\Omega} |tg| \, \mathrm{d}(s,r)$$

Note that this theorem also implies the continuity of the norm on $L^{p,q}(\Sigma \times \Omega)$ with respect to the strong topology.

Remark 2.8. With Theorem 2.7 we can directly infer that on $L^{p,q}(\Sigma \times \Omega)$ an analogue of the Minkowski inequality applies, i.e., for arbitrary functions $t, h \in L^{p,q}(\Sigma \times \Omega)$ we have

$$\|t + h\|_{L^{p,q}(\Sigma\times\Omega)} \leq \|t\|_{L^{p,q}(\Sigma\times\Omega)} + \|h\|_{L^{p,q}(\Sigma\times\Omega)}.$$

Therefore, we deduce that

$$\begin{aligned} \|t - h + h\|_{L^{p,q}} \leq \|t - h\|_{L^{p,q}} + \|h\|_{L^{p,q}} \;&\Leftrightarrow\; \|t\|_{L^{p,q}} - \|h\|_{L^{p,q}} \leq \|t - h\|_{L^{p,q}}, \\ \|t - h - t\|_{L^{p,q}} \leq \|t - h\|_{L^{p,q}} + \|t\|_{L^{p,q}} \;&\Leftrightarrow\; \|h\|_{L^{p,q}} - \|t\|_{L^{p,q}} \leq \|t - h\|_{L^{p,q}}, \end{aligned}$$

and conclude that for arbitrary $\varepsilon > 0$ and functions $t, h \in L^{p,q}(\Sigma \times \Omega)$ with $\|t-h\|_{L^{p,q}(\Sigma\times\Omega)} < \varepsilon$ it can be guaranteed that

$$\left| \, \|t\|_{L^{p,q}(\Sigma\times\Omega)} - \|h\|_{L^{p,q}(\Sigma\times\Omega)} \right| < \varepsilon.$$

Thus, regarding the strong topology the norm on $L^{p,q}(\Sigma \times \Omega)$ is continuous.

Speaking of the strong topology, we do not want to neglect mentioning the behavior of $L^{p,q}(\Sigma\times\Omega)$ regarding the weak topology.

Theorem 2.9. (cf. [Benedek and Panzone, 1961, Sec. 5, Thrm. 2])
Let $1 \leq p, q \leq \infty$. Then, $L^{p,q}(\Sigma \times \Omega)$ is weakly sequentially complete.

Notice that with this statement it is moreover clear that, regarding the weak topology, the norm on $L^{p,q}(\Sigma \times \Omega)$ obeys a lower semicontinuity.

Remark 2.10. Let $(t_\gamma)_{\gamma \in \mathbb{N}} \subset L^{p,q}(\Sigma \times \Omega)$ be a weakly convergent sequence with limit t_*. Then first of all, due to Theorem 2.9, $t_* \in L^{p,q}(\Sigma \times \Omega)$. Second, we can deduce with Hölder's inequality that for all $g \in L^{p',q'}(\Sigma \times \Omega)$ with $\|g\|_{L^{p',q'}(\Sigma \times \Omega)} = 1$

$$\int_\Sigma \int_\Omega |t_\gamma g| \, \mathrm{d}r \, \mathrm{d}s \leq \|t_\gamma\|_{L^{p,q}(\Sigma \times \Omega)}$$

applies. Hence, we realize that

$$\left| \int_{\Sigma \times \Omega} t_* g \, \mathrm{d}(s,r) \right| \leq \liminf_{\gamma \to \infty} \int_{\Sigma \times \Omega} |t_\gamma g| \, \mathrm{d}(s,r) \leq \liminf_{\gamma \to \infty} \|t_\gamma\|_{L^{p,q}(\Sigma \times \Omega)}$$

and conclude together with Theorem 2.7 that

$$\|t_*\|_{L^{p,q}(\Sigma \times \Omega)} = \sup_{\|g\|_{L^{p',q'}(\Sigma \times \Omega)} = 1} \left| \int_{\Sigma \times \Omega} t_* g \, \mathrm{d}(s,r) \right| \leq \liminf_{\gamma \to \infty} \|t_\gamma\|_{L^{p,q}(\Sigma \times \Omega)},$$

i.e., that the norm on $L^{p,q}(\Sigma \times \Omega)$ is lower semicontinuous with respect to the weak topology.

To conclude this section on function spaces, the last useful statement we want to mention, again, deals with the density of the space of C_c^∞ functions.

Lemma 2.11. Let $1 \leq p, q < \infty$. Then, $C_c^\infty(\Sigma \times \Omega)$ lies dense in $L^{p,q}(\Sigma \times \Omega)$.

Proof. Let $t \in L^{p,q}(\Sigma \times \Omega)$. Then, we realize that

$$\|t\|_{L^{p,q}(\Sigma \times \Omega)} = \left(\int_\Sigma \left(\int_\Omega |t(s,r)|^q \, \mathrm{d}r \right)^{\frac{p}{q}} \mathrm{d}s \right)^{\frac{1}{p}} = \left(\int_\Sigma \|t(s,\,\cdot\,)\|_{L^q(\Omega)}^p \, \mathrm{d}s \right)^{\frac{1}{p}},$$

i.e., strictly speaking $L^{p,q}(\Sigma \times \Omega)$ can be identified with the Bochner space $L^p(\Sigma; L^q(\Omega))$ (cf., e.g., [Hytönen et al., 2016, Def. 1.2.15]). Now, respecting that for these Bochner spaces it applies that $C_c^\infty(\Sigma; \mathcal{Y})$ lies dense in $L^p(\Sigma; \mathcal{Y})$ for $\mathcal{Y}$ being a Banach space (cf., e.g., [Hytönen et al., 2016, Lem. 1.2.31]), and involving that $C_c^\infty(\Omega)$ also lies dense in $L^q(\Omega)$, we can again follow with standard arguments that $C_c^\infty(\Sigma \times \Omega)$ is dense in $L^{p,q}(\Sigma \times \Omega)$. □

2.2 Integral operators

The next concept we want to recall and which is highly correlated with the just introduced mixed Lebesgue spaces is the one which deals with integral operators and their singular value decomposition.

Considering open subsets $\Sigma \subset \mathbb{R}^d$ and $\Omega \subset \mathbb{R}^{d'}$ and a function $t \in L^{p,q}(\Sigma \times \Omega)$, within this work we are predominantly interested in contemplating *integral operators* $T : L^{q'}(\Omega) \to L^p(\Sigma)$ characterized through

$$(Ty)(s) = \int_\Omega t(s,r)y(r)\,\mathrm{d}r \quad \forall s \in \Sigma. \tag{2.2}$$

In this setting we want to call t the *integral kernel* associated with T. Note that, owing to Hölder's inequality and

$$\begin{aligned} ||Ty||^p_{L^p(\Sigma)} &= \int_\Sigma \left| \int_\Omega t(s,r)y(r)\,\mathrm{d}r \right|^p \mathrm{d}s \\ &\leq \int_\Sigma \left(\int_\Omega |t(s,r)|^q\,\mathrm{d}r \right)^{\frac{p}{q}} \left(\int_\Omega |y(r)|^{q'}\,\mathrm{d}r \right)^{\frac{p}{q'}} \mathrm{d}s \\ &= ||t||^p_{L^{p,q}(\Sigma\times\Omega)}\, ||y||^p_{L^{q'}(\Omega)}, \end{aligned}$$

operators of this form are linear and bounded and, hence, the convergence of sequences of integral operators can be understood in the classical sense of operator topologies given to the space $L(\mathcal{Y}, \mathcal{X})$.

Definition 2.12. (cf. [Riesz and Sz-Nagy, 1955, §66])
Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces and $(T_\gamma)_{\gamma\in\mathbb{N}} \subset L(\mathcal{Y}, \mathcal{X})$ be a sequence of operators. Then,

(i) $(T_\gamma)_{\gamma\in N}$ converges in the *uniform operator topology* toward an operator $T : \mathcal{Y} \to \mathcal{X}$ if

$$\lim_{\gamma\to\infty} \|T_\gamma - T\| = 0,$$

(ii) $(T_\gamma)_{\gamma\in\mathbb{N}}$ converges in the *strong operator topology* toward an operator $T : \mathcal{Y} \to \mathcal{X}$ if

$$\lim_{\gamma\to\infty} \|T_\gamma y - Ty\|_{\mathcal{X}} = 0 \quad \forall y \in \mathcal{Y},$$

(iii) $(T_\gamma)_{\gamma\in\mathbb{N}}$ converges in the *weak operator topology* toward an operator $T : \mathcal{Y} \to \mathcal{X}$ if

$$\lim_{\gamma\to\infty} |f(T_\gamma y) - f(Ty)| = 0 \quad \forall y \in \mathcal{Y}, \forall f \in \mathcal{X}^*.$$

2.2.1 Singular value decomposition

As we will see in Chapter 6, in this work integral operators of the form (2.2) will become especially interesting as soon as a so-called singular system can be assigned to them. In order to enable this, these operators first of all have to be compact.

Definition 2.13. (cf., e.g., [Brezis, 2010, Def. 6.1])
Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces. A linear operator $T : \mathcal{Y} \to \mathcal{X}$ is said to be *compact* if for every bounded set $\mathcal{C} \subset \mathcal{Y}$ the image $T(\mathcal{C})$ has a compact closure in $\mathcal{X}$.

Notice that compact operators own the unique feature to be able to improve the characteristic of the convergence of a sequence.

Lemma 2.14. (cf., e.g., [Brezis, 2010, Rem. 6.2])
Let $T \in L(\mathcal{Y}, \mathcal{X})$ be compact. Suppose that $(y_\gamma)_{\gamma\in\mathbb{N}}$ converges weakly to y in $\mathcal{Y}$. Then, $(Ty_\gamma)_{\gamma\in\mathbb{N}}$ converges strongly toward Ty in $\mathcal{X}$.

Coming back to our integral operators we perceive that especially those with kernel $t \in L^2(\Sigma\times\Omega)$ satisfy the requirements of compactness. This can be realized when approximating t with a degenerate kernel, i.e.,

$$t(s,r) \approx t_n(s,r) := \sum_{j=1}^{n} g_j(s)h_j(r)$$

for $g_j(s) := \int_{\Omega_j^n} t(s,r)\,\mathrm{d}r \,/\, |\Omega_j^n|$, $h_j(r) := \chi_{\Omega_j^n}(r)$ with $\Omega = \dot{\bigcup}_{j=1}^{n}\Omega_j^n$ and $|\Omega_j^n| \le \frac{C}{n}$. Considering the corresponding sequence of integral operators $(T_n)_{n\in\mathbb{N}}$ with kernels $(t_n)_{n\in\mathbb{N}}$ we can observe that all of its elements have a finite-dimensional range and thus, due to their boundedness, are compact. With that, we have found a sequence of compact operators which, with respect to the uniform operator topology, converges to T and therefore implies its compactness (cf. [Brezis, 2010, Cor. 6.2]). As a result, we can state that, from this point of view, integral operators with kernel $t \in L^2(\Sigma \times \Omega)$ qualify for the assignment of a singular system. And in fact, since the spaces $L^2(\Sigma)$ and $L^2(\Omega)$ also represent Hilbert spaces this suitability can be confirmed.

Definition 2.15. (cf., e.g., [Engl, 1997, Def. 7.13])
Let $\mathcal{X}$ and $\mathcal{Y}$ be Hilbert spaces and $T \in L(\mathcal{Y}, \mathcal{X})$ be compact. A sequence $(\sigma_\xi, u_\xi, v_\xi)_{\xi\in\mathbb{N}}$ is called *singular system* of T, if: $\sigma_\xi > 0$ for all $\xi \in \mathbb{N}$, $(\sigma_\xi^2, u_\xi)_{\xi\in\mathbb{N}}$ is an eigensystem to TT^* and $v_\xi = \frac{T^*u_\xi}{\|T^*u_\xi\|}$ for all $\xi \in \mathbb{N}$.

Note that within the course of this work we will sometimes also make use of the notation $(\sigma_\xi(T))_{\xi\in\mathbb{N}}$ in order to explicitly indicate that the contemplated singular values correspond to the operator T. Besides, we want to point to the following proposition which illustrates the characteristic of a singular system in more detail.

Proposition 2.16. (cf. [Engl, 1997, Prop. 7.14])
Let $T \in L(\mathcal{Y}, \mathcal{X})$ be a compact operator with associated singular system $(\sigma_\xi, u_\xi, v_\xi)_{\xi\in\mathbb{N}}$. Then, the following applies:

(i) For all $\xi \in \mathbb{N}$ it holds that

$$T^* u_\xi = \sigma_\xi v_\xi, \quad T v_\xi = \sigma_\xi u_\xi,$$

(ii) $(\sigma_\xi^2, v_\xi)_{\xi \in \mathbb{N}}$ represents an eigensystem to T^*T,

(iii) $(u_\xi)_{\xi \in \mathbb{N}}$ and $(v_\xi)_{\xi \in \mathbb{N}}$ are orthonormal bases of $\overline{\operatorname{ran} T}$, respectively of $\overline{\operatorname{ran} T^*} = \ker T^\perp$.

Proof. Let $T \in L(\mathcal{Y}, \mathcal{X})$ with corresponding singular system $(\sigma_\xi, u_\xi, v_\xi)_{\xi \in \mathbb{N}}$. Then, according to Definition 2.15 it holds that $TT^* u_\xi = \sigma_\xi^2 u_\xi$ for all $\xi \in \mathbb{N}$. Hence, also $T^* u_\xi \neq 0$ applies and v_ξ is well-defined. Moreover, we can compute that

$$\begin{aligned} T v_\xi &= \frac{1}{\sqrt{\langle T^* u_\xi, T^* u_\xi \rangle}} TT^* u_\xi = \frac{\sigma_\xi^2 u_\xi}{\sqrt{\langle u_\xi, TT^* u_\xi \rangle}} = \frac{\sigma_\xi^2 u_\xi}{\sqrt{\sigma_\xi^2 \|u_\xi\|^2}} = \sigma_\xi u_\xi, \\ T^* u_\xi &= v_\xi \|T^* u_\xi\| = v_\xi \sqrt{\langle u_\xi, TT^* u_\xi \rangle} = \sigma_\xi v_\xi. \end{aligned}$$

This, in particular, implies that

$$T^* T v_\xi = \sigma_\xi T^* u_\xi = \sigma_\xi^2 v_\xi,$$

and we can identify $(\sigma_\xi^2, v_\xi)_{\xi \in \mathbb{N}}$ to represent an eigensystem to T^*T. With that, since eigenvectors of compact selfadjoint operators always form a orthonormal basis of the closure of its range (cf. [Engl, 1997, Prop. 2.38]), we can be sure that $(u_\xi)_{\xi \in \mathbb{N}}$ and $(v_\xi)_{\xi \in \mathbb{N}}$ represent orthonormal bases of $\overline{\operatorname{ran} TT^*}$, respectively of $\overline{\operatorname{ran} T^*T}$. Now, on the one hand

$$\operatorname{ran} T = T(\ker T^\perp) = T(\overline{\operatorname{ran} T^*}) \subseteq \overline{T(\operatorname{ran} T^*)} = \overline{\operatorname{ran} TT^*}$$

and analogously $\operatorname{ran} T^* \subseteq \overline{\operatorname{ran} T^*T}$ applies. However, on the other hand obviously also $\operatorname{ran} TT^* \subseteq \operatorname{ran} T$ and $\operatorname{ran} T^*T \subseteq \operatorname{ran} T^*$ holds true. Consequently, $\overline{\operatorname{ran} T} = \overline{\operatorname{ran} TT^*}$ and $\overline{\operatorname{ran} T^*} = \overline{\operatorname{ran} T^*T}$ and the proposition is proven. □

In the preceding paragraph we have thus learned that for $p = q = 2$ the integral operators in (2.2) can be equipped with a singular system. Now, to conclude this section, we briefly want to comment on the impact this system has on their representation.

Lemma 2.17. Let $T : L^2(\Omega) \to L^2(\Sigma)$ be an integral operator and $(\sigma_\xi, u_\xi, v_\xi)_{\xi \in \mathbb{N}}$ denote an associated singular system. Then, the inducing integral kernel $t \in L^2(\Sigma \times \Omega)$ can be described by

$$t(s, r) = \sum_{\xi \in \mathbb{N}} \sigma_\xi u_\xi(s) v_\xi(r) \quad \forall s \in \Sigma, r \in \Omega.$$

Proof. Suppose that $T : L^2(\Omega) \to L^2(\Sigma)$ is an integral operator and $(\sigma_\xi, u_\xi, v_\xi)_{\xi\in\mathbb{N}}$ denotes a corresponding singular system. Then $(w_{\phi\psi})_{\phi,\psi\in\mathbb{N}}$ characterized by

$$w_{\phi\psi}(s,t) \coloneqq u_\phi(s)v_\psi(r) \quad \forall s \in \Sigma, r \in \Omega$$

represents a system in $L^2(\Sigma \times \Omega)$ which in consequence of

$$\begin{aligned}\langle w_{\phi\psi}, w_{\phi'\psi'}\rangle &= \int_\Sigma \int_\Omega u_\phi(s)v_\psi(r)\, u_{\phi'}(s)v_{\psi'}(r)\,\mathrm{d}r\,\mathrm{d}s \\ &= \int_\Sigma u_\phi(s)u_{\phi'}(s)\,\mathrm{d}s \int_\Omega v_\psi(r)v_{\psi'}(r)\,\mathrm{d}r = \delta_{\phi\phi'}\delta_{\psi\psi'}\end{aligned}$$

is orthonormal. Perceiving that for all $x \in \operatorname{ran} T^\perp$, $y \in L^2(\Omega)$

$$\langle t, xy\rangle = \int_\Sigma \int_\Omega t(s,r)\, x(s)y(r)\,\mathrm{d}r\,\mathrm{d}s = \int_\Sigma x(s) \int_\Omega t(s,r)y(r)\,\mathrm{d}r\,\mathrm{d}s = \langle x, Ty\rangle = 0$$

and for all $x \in L^2(\Sigma)$, $y \in \operatorname{ran} T^{*\perp}$

$$\langle t, xy\rangle = \langle T^* x, y\rangle = 0$$

applies we can deduce the belonging of t to the subspace of $L^2(\Sigma \times \Omega)$ which is spanned by $(w_{\phi\psi})_{\phi,\psi\in\mathbb{N}}$. Hence, t can be described by

$$t = \sum_{\phi\in\mathbb{N}} \sum_{\psi\in\mathbb{N}} \langle t, w_{\phi\psi}\rangle\, w_{\phi\psi}$$

and we conclude together with

$$\langle t, w_{\phi\psi}\rangle = \int_\Sigma \int_\Omega t(s,r)\, u_\phi(s)v_\psi(r)\,\mathrm{d}r\,\mathrm{d}s = \int_\Sigma u_\phi(s)(Tv_\psi)(s)\,\mathrm{d}s = \int_\Sigma u_\phi\sigma_\psi u_\psi(s)\,\mathrm{d}s = \sigma_\psi\delta_{\phi\psi}$$

that

$$t(s,r) = \sum_{\phi\in\mathbb{N}} \sum_{\psi\in\mathbb{N}} \sigma_\psi\delta_{\phi\psi} w_{\phi\psi}(s,r) = \sum_{\psi\in\mathbb{N}} \sigma_\psi u_\psi(s)v_\psi(r) \quad \forall s \in \Sigma, r \in \Omega.$$

□

Note that through this representation an integral operator $T : L^2(\Omega) \to L^2(\Sigma)$ with associated singular system $(\sigma_\xi, u_\xi, v_\xi)_{\xi\in\mathbb{N}}$ can generally be described via

$$T = \sum_{\xi\in\mathbb{N}} \sigma_\xi (v_\xi \otimes u_\xi).$$

2.3 Γ-convergence

In the last section of this preliminary chapter we want to focus on the so-called Γ-*convergence* of sequences of functionals.
As we will see in the analytical part of this work, when dealing with sequences of functionals, in a natural way often the question of their limit behavior arises. In Definition 2.12 we have already seen how the convergence of bounded and linear operators can be understood. However, especially in the scenario in which one is predominantly interested in the minimizers of the elements of the contemplated sequence, it may be more suitable to apply a different concept. Focusing on these minimizers it would be desirable to be able to consider a notion of convergence which preserves the minimizing structure of the sequence in the sense that the induced sequence of minimizers converges toward the minimizer of the assigned limit functional. To implement this idea in the 1970s De Giorgi introduced the concept of Γ-convergence.

Definition 2.18. (cf., e.g., [Braides, 2002, Def. 1.5])
Let $\mathcal{X}$ be a metric space and $(\mathcal{F}_\mu)_{\mu\in\mathbb{N}}$ with $\mathcal{F}_\mu : \mathcal{X} \to \overline{\mathbb{R}}$ be a sequence of functionals. Then, $(\mathcal{F}_\mu)_{\mu\in\mathbb{N}}$ is said to Γ-*converge* in $\mathcal{X}$ to $\mathcal{F} : \mathcal{X} \to \overline{\mathbb{R}}$ if for all $x \in \mathcal{X}$ the following applies:

(i) *lim inf-inequality:* For every sequence $(x_\mu)_{\mu\in\mathbb{N}}$ converging to x it holds that

$$\mathcal{F}(x) \le \liminf_{\mu\to\infty} \mathcal{F}_\mu(x_\mu),$$

(ii) *Existence of a recovery sequence:* There exists a sequence $(x_\mu)_{\mu\in\mathbb{N}}$ converging to x such that

$$\mathcal{F}(x) = \lim_{\mu\to\infty} \mathcal{F}_\mu(x_\mu).$$

Nevertheless, in order to pursue the above-mentioned goal and obtain a limit functional which is compatible with the minimizing structure of the sequence, it does not suffice to determine the Γ-limit. In addition, the following property has to apply.

Definition 2.19. (cf., e.g., [Dal Maso, 1993, Def. 7.6])
A sequence of functionals $(\mathcal{F}_\mu)_{\mu\in\mathbb{N}}$ on $\mathcal{X}$ is called *equi-coercive* if for all sequences $(x_\mu)_{\mu\in\mathbb{N}} \subset \mathcal{X}$, which fulfill

$$\sup_{\mu\in\mathbb{N}} \mathcal{F}_\mu(x_\mu) < \infty,$$

it holds, up to subsequences, that they converge toward some $x \in \mathcal{X}$.

Combining the characteristics of both of these definitions then, in fact, leads to the desired convergence behavior.

Theorem 2.20. (cf., e.g., [Braides, 2002, Thrm. 1.21])
Let $(\mathcal{F}_\mu)_{\mu\in\mathbb{N}}$ be an equi-coercive sequence of functionals on $\mathcal{X}$ which Γ-converges to $\mathcal{F}$. Then,

$$\min_{x\in\mathcal{X}} \mathcal{F}(x) = \lim_{\mu\to\infty} \inf_{x\in\mathcal{X}} \mathcal{F}_\mu(x).$$

Moreover, if $(x_\mu)_{\mu\in\mathbb{N}}$ is a precompact sequence such that

$$\lim_{\mu\to\infty} \mathcal{F}_\mu(x_\mu) = \lim_{\mu\to\infty} \inf_{x\in\mathcal{X}} \mathcal{F}_\mu(x),$$

then every limit of a subsequence of $(x_\mu)_{\mu\in\mathbb{N}}$ is a minimum point for $\mathcal{F}$.

Thus, under the assumption of equi-coercivity and Γ-convergence, it can be guaranteed that minimizers converge toward minimizers. However, note that in Definitions 2.18 and 2.19 we did not specify the topology on $\mathcal{X}$ with respect to which this convergence is to be understood. This, in general, leaves room for interpretation and allows to adapt the assertion in Theorem 2.20 to various settings. Though, in order to apply the preceding result, one always has to make sure that the topologies considered to show equi-coercivity and Γ-convergence coincide. Therefore, when aiming for a structure-preserving limit, it is advisable to first of all identify the (strongest) topology rendering the sequence of functionals equi-coercive, before then tackling the proof of Γ-convergence with respect to this very topology.
A further advice, which in practice often facilitates proving the Γ-convergence of a sequence of functionals, is concerned with the existence of a recovery sequence.

Remark 2.21. (cf., e.g., [Braides, 2002, Rem. 1.29])
Let the upper Γ-limit of a sequence of functionals $(\mathcal{F}_\mu)_{\mu\in\mathbb{N}}$ with $\mathcal{F}_\mu : \mathcal{X} \to \overline{\mathbb{R}}$ be defined as

$$\Gamma\text{-}\limsup_\mu \mathcal{F}_\mu(x) \coloneqq \inf \left\{ \limsup_{\mu\to\infty} \mathcal{F}_\mu(x_\mu) \,\middle|\, (x_\mu)_{\mu\in\mathbb{N}} \text{ with } x_\mu \xrightarrow{d} x \in \mathcal{X} \right\},$$

where $\xrightarrow{d}$ denotes the convergence with respect to the topology induced by a metric d. Let furthermore $\mathcal{D} \subseteq \mathcal{X}$ lie dense in $\mathcal{X}$ with respect to the topology induced by a metric d' which is not weaker than the one regarding d, i.e.,

$$\forall x \in \mathcal{X} \quad \exists\, (x_\gamma)_{\gamma\in\mathbb{N}} \subset \mathcal{D} : \quad x_\gamma \xrightarrow{d'} x.$$

Then, since it can be shown that $\Gamma\text{-}\limsup_\mu \mathcal{F}_\mu$ is a lower semicontinuous function (cf., e.g., [Braides, 2002, Prop. 1.28]), proving that

$$\Gamma\text{-}\limsup_\mu \mathcal{F}_\mu \leq \mathcal{F} \quad \text{on } \mathcal{D}$$

implies that for any $x \in \mathcal{X}$ there exists a sequence $(x_\gamma)_{\gamma\in\mathbb{N}} \subset \mathcal{D}$ with $x_\gamma \xrightarrow{d'} x$ and

$$\Gamma\text{-}\limsup_\mu \mathcal{F}_\mu(x) \leq \liminf_{\gamma\to\infty} \left(\Gamma\text{-}\limsup_\mu \mathcal{F}_\mu(x_\gamma)\right) \leq \liminf_{\gamma\to\infty} \mathcal{F}(x_\gamma).$$

Now, supposing that $\mathcal{F}$ was continuous with respect to the topology induced by d, which then obviously transfers to the one regarding d', we obtain that

$$\Gamma\text{-}\limsup_{\mu} \mathcal{F}_\mu \leq \mathcal{F} \quad \text{on } \mathcal{X}.$$

Assuming that the lim inf-inequality was already shown, this means that

$$\left[\forall x \in \mathcal{D}\ \exists\, (x_\mu)_{\mu\in\mathbb{N}} :\ \mathcal{F}(x) = \lim_{\mu\to\infty} \mathcal{F}_\mu(x_\mu)\right] \ \Leftrightarrow\ \left[\forall x \in \mathcal{X}\ \exists\, (x_\mu)_{\mu\in\mathbb{N}} :\ \mathcal{F}(x) = \lim_{\mu\to\infty} \mathcal{F}_\mu(x_\mu)\right],$$

and we have found that it suffices to prove the existence of a recovery sequence on a dense subset of the original domain $\mathcal{X}$.

Hence, with this insight one can exploit, that in some scenarios it may be easier to switch to a dense subset when facing the second requirement in Definition 2.18.
Beyond that, when contemplating sequences of composed functionals, it can be beneficial to be aware of the existence of another different notion of convergence.

Definition 2.22. (cf., e.g., [Dal Maso, 1993, Def. 4.7])
A sequence of functionals $(\mathcal{F}_\mu)_{\mu\in\mathbb{N}}$ on $\mathcal{X}$ is said to be *continuously convergent* (in $\mathcal{X}$) to a function $\mathcal{F} : \mathcal{X} \to \overline{\mathbb{R}}$ if for all sequences $(x_\mu)_{\mu\in\mathbb{N}} \subset \mathcal{X}$ with limit x

$$\lim_{\mu\to\infty} \mathcal{F}_\mu(x_\mu) = \mathcal{F}(x)$$

applies.

Obviously, this type of convergence is stronger than the previously defined Γ-convergence. Therefore, when combining two sequences of functionals, one of them converging in the Γ-sense and the other in the continuous sense, the logical consequence on their joint convergence behavior seems to be the following.

Proposition 2.23. (cf. [Dal Maso, 1993, Prop. 6.20])
Let $(\mathcal{E}_\mu)_{\mu\in\mathbb{N}}$ and $(\mathcal{F}_\mu)_{\mu\in\mathbb{N}}$ be sequences of functionals on $\mathcal{X}$. Suppose that $(\mathcal{E}_\mu)_{\mu\in\mathbb{N}}$ is continuously convergent to a function $\mathcal{E}$ and that $\mathcal{E}_\mu$ and $\mathcal{E}$ are everywhere finite on $\mathcal{X}$. Then, if $(\mathcal{F}_\mu)_{\mu\in\mathbb{N}}$ Γ-converges to $\mathcal{F}$ in $\mathcal{X}$, $(\mathcal{E}_\mu + \mathcal{F}_\mu)_{\mu\in\mathbb{N}}$ Γ-converges to $\mathcal{E} + \mathcal{F}$ in $\mathcal{X}$.

Consequently, when considering sequences of such composed functionals and perceiving that the sequence of one of their components is continuously convergent, with respect to the determination of a Γ-limit one can neglect this part and concentrate on the remainder.

With this, we are now equipped with all technical tools that allow to perform the analysis in Chapter 6. However, to be able to also follow the preceding chapters, in addition it is necessary to familiarize with the field of inverse problems. This is what we want to look into in the following chapter.

3

Inverse problems and variational methods

In the previous chapter we gave a quick overview of a few mathematical concepts that will play a relevant role within the scope of this thesis. However, up to now we omitted to properly introduce the general theory on which this work is based: the theory of inverse problems.
To do so we briefly want to recall the general setting inverse problems are located in and point to the associated difficulties regarding their solvability. Discussing some approaches to eliminate these inconveniences then leads us to the introduction of the concept of regularization. Building on this we involve the Bayesian modeling of corrupted data and shortly present how to tackle inverse problems via the application of variational methods.
Since the demonstrations in this chapter will be kept on a fundamental level, we refer the reader to [Kirsch, 2011] and [Brinkmann, 2019, Chap. 4] for a more detailed discussion.

3.1 Inverse problems

In mathematical imaging an *inverse problem* is generally understood to be the hunt after the origin of observed consequences generated by a specified model. In more formal words this means when considering an operator K between Banach spaces $\mathcal{Y}$ and $\mathcal{X}$ one is interested in finding the entity $y \in \mathcal{Y}$ which through

$$Ky = x^{\Diamond} \tag{3.1}$$

caused the observation of the given element $x^{\Diamond} \in \mathcal{X}$. Throughout this work we want to focus on linear and bounded operators, i.e., we exclusively contemplate $K \in L(\mathcal{Y}, \mathcal{X})$. For problems of this kind, in 1902, Hadamard developed a guideline to define their so-called *well-posedness* – a measure for their exact and reasonable solvability.

Definition 3.1. (cf. [Hadamard, 1902])
Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces and $K : \mathcal{Y} \to \mathcal{X}$ define a linear and bounded operator. Then, problems of the form (3.1) are called *well-posed* as soon as the subsequent conditions are satisfied for all $x^{\Diamond} \in \mathcal{X}$:

(i) *Existence:* there exists a solution $y \in \mathcal{Y}$ such that $Ky = x^\lozenge$ applies,

(ii) *Uniqueness:* the solution y is unique,

(iii) *Stability:* the solution y depends continuously on the input data.

If at least one of these requirements is violated the problem is called *ill-posed.*

Thus, in order to be able to provide proper solutions, problems of the form (3.1) need to involve surjective and injective, hence bijective, operators K whose inverse operators K^{-1} are continuous. Accordingly, regarding the first two conditions, two main problems may arise: First of all it may occur that $\operatorname{ran} K \subsetneq \mathcal{X}$, second the operator K may be given such that $\ker K \neq \{0\}$. However, these can be remedied quite easily (cf., e.g., [Engl et al., 1996]). By engaging with an approximate solution one can agree on introducing the *least squares solution* of (3.1). Defined as the element $y \in \mathcal{Y}$ which minimizes

$$||Ky - x^\lozenge||_\mathcal{X}$$

it extends the solution set and therefore allows for a more generalized understanding of solvability. To prevent the occurrence of multiple least squares solutions, which would still violate Hadamard's second demand, one can then add a supplementary claim. A very popular one is the claim for a small norm. Applied to our problem this means that among all least squares solutions the one with the smallest norm is selected. The resulting element is then commonly called the *minimum norm solution.* In order to implement both of these ideas simultaneously, the definition of a generalized inverse was established.

Definition 3.2. (cf., e.g., [Engl et al., 1996, Def. 2.2])
Let $K \in L(\mathcal{Y}, \mathcal{X})$ and $\widetilde{K}$ be characterized through $\widetilde{K} := K|_{\ker K^\perp} : \ker K^\perp \to \operatorname{ran} K$. Then, the *Moore-Penrose inverse* $K^\dagger$ of K is defined as the unique linear extension of $\widetilde{K}^{-1}$ to $\operatorname{dom} K^\dagger := \operatorname{ran} K \oplus \operatorname{ran} K^\perp$ with $\ker K^\dagger = \operatorname{ran} K^\perp$.

In fact, it can be shown (see, e.g., [Engl et al., 1996, Thrm. 2.5]) that as soon as $x^\lozenge \in \operatorname{dom} K^\dagger$ the minimum norm solution to (3.1) is given by $K^\dagger x^\lozenge$. Consequently, the Moore-Penrose inverse seems to be a promising tool to avoid issues regarding the well-posedness of inverse problems. Nevertheless, during its construction we neglected the stability problem. This usually stems from the fact, that most of the considered operators of interest are compact. Recalling the characterization in Definition 2.13 we realize that all linear and bounded operators K mapping to a finite-dimensional Banach space $\mathcal{X}$ are directly affected by this. In addition, however, also other relevant operators suffer from this compactness. For example, it can be demonstrated that this includes all integral operators defined on L^2-spaces (see, e.g., [Engl, 1997, Prop. 2.11]). The explicit problem arising from this property becomes clear in the following theorem, which deems to be a direct consequence from Riesz' Lemma (see, e.g., [Rudin, 1991, Lem. 4.22]).

Theorem 3.3. (cf., e.g., [Lebedev et al., 2003, Thrm. 6.5.4])
Let $\mathcal{Y}$ be an infinite-dimensional Banach space and suppose that $K \in L(\mathcal{Y}, \mathcal{X})$ is compact. Then, K can not posses a continuous inverse.

Hence, the consideration of models involving compact operators violates the third condition of well-posedness. Unfortunately, it is quite easy to prove (see, e.g., [Engl et al., 1996, Prop. 2.7]) that the toxic property in Theorem 3.3 also transfers to the previously introduced Moore-Penrose inverse $K^{\dagger}$. Thus, also through this alternative solution approach the just described problem can not be circumvented.
Beyond that, even when considering models which involve non-compact operators, there exists a previously concealed feature which restricts the use of the Moore-Penrose inverse: Although this pseudo inverse is defined on a larger domain than $\operatorname{ran} K$ it still does not necessarily cover all elements in $\mathcal{X}$. A full coverage can only be reached if the range of K is closed, otherwise there still exist elements $x^{\Diamond} \in \mathcal{X}$ for which no minimum norm solution can be found.
With respect to real world problems these two limitations on the use of the Moore-Penrose inverse represent major drawbacks. This is because, as already insinuated within the introduction of this work, due to not entirely exact measurements the generated data of most imaging-related applications is noise-affected. Consequently, it is first of all possible that, in the case of non-closed $\operatorname{ran} K$, these data do not lie in the domain of $K^{\dagger}$ and no solution can be determined. On the other hand it is likely that, in the case of discontinuous $K^{\dagger}$, its application will yield a solution which is not proportional to the deviation of the noisy data from the exact data, and thus unusable.
Accordingly, we quickly understand that we have to use other methods to ensure solvability and the desired stability. In order to address these, from now on we want to consider problems of the form

$$Ky = x. \tag{3.2}$$

Here $x \in \mathcal{X}$ represents a version of the exact data $x^{\Diamond}$ which is noisy to the level $\delta > 0$, i.e., for which $||x - x^{\Diamond}||_{\mathcal{X}} \leq \delta$ applies. To meet the goal of stability, in this setting we would then like to accept solutions $y \in \mathcal{Y}$ that fulfill

$$||Ky - x||_{\mathcal{X}} \leq \delta. \tag{3.3}$$

A very intuitive idea for solving (3.2) was provided by Tikhonov in the 1960s (see [Tikhonov and Arsenin, 1977] and references therein). It proposed to identify the minimizer of the norm of the residual under the condition of the boundedness of its own norm, i.e., to solve

$$\arg\min_{y \in \mathcal{Y}} ||Ky - x||_{\mathcal{X}} \quad \text{subject to} \quad ||y||_{\mathcal{Y}} \leq r$$

for a radius $r > 0$. Respecting the monotonicity of the norm and its squared version as well as introducing a Lagrange multiplier α depending on r, this approach can be translated to solving

$$\arg\min_{y\in\mathcal{Y}} \frac{1}{2}||Ky - x||^2_{\mathcal{X}} + \frac{\alpha}{2}||y||^2_{\mathcal{Y}}. \tag{3.4}$$

Here it is easy to see, that for small α the solution of this minimization problem is indeed a good estimate for the solution of (3.2). However, in order to meet the demand in (3.3) the exact choice of this parameter should obviously depend on the given noise level δ.
With this proposal, similar to the concept of the minimum norm solution, Tikhonov's idea was to put requirements on the norm of the residual as well as on the norm of the sought-after object in order to guarantee the existence of a unique solution. Beyond that, due to the constraint of boundedness on the norm of y, his approach was able to elude the potential discontinuity the Moore-Penrose inverse was suffering from. Hence, Tikhonov had developed a strategy to approximate ill-posed inverse problems by well-posed ones and with this introduced the concept of *regularization*.

3.2 Variational methods

Through Tikhonov's approach in (3.4) we have already got an idea how ill-posed inverse problems can be solved approximately. Here, in order to ensure that the inexact result only deviates in a reasonable manner from an exact one, we asked the norm of the corresponding residual to be minimal. Simultaneously, to differentiate between optional solutions which fulfill the desired closeness in a comparable manner and to bound the set of solutions, we resorted to the claim of a minimal norm of the solution itself.
Although, in the light of the aforementioned, the objectives pursued by these two criteria seem plausible, their exact choice appears to be a little bit random. Why should one consider the norm of the residual as a measure for how well (3.2) is fulfilled? And why should a solution with minimal natural norm be eligible?
These are the questions that *variational modeling* is trying to answer. Allowing for alternative approaches to pursue the just mentioned objectives, it concentrates on the design of non-generic minimization problems which adapt to the specific circumstances and conditions which surround the respective inverse problems. Mathematically speaking this means that variational modeling has made it its business to formulate suitable problems of the form

$$\arg\min_{y\in\mathcal{Y}} \mathcal{F}_\alpha(y), \tag{3.5}$$

where $\mathcal{F}_\alpha : \mathcal{Y} \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ represents an energy functional which is composed of a functional $\mathcal{D}_x : \mathcal{X} \to \mathbb{R}$ measuring the data fidelity and a regularizing one $\mathcal{R} : \mathcal{Y} \to \overline{\mathbb{R}}$, i.e.,

$$\mathcal{F}_\alpha(y) := \mathcal{D}_x(Ky) + \alpha\mathcal{R}(y).$$

In order to grasp what exactly is meant when speaking about conditions and circumstances and to get a sense for how suitable choices for $\mathcal{D}_x$ and $\mathcal{R}$ could look like, in the following we want to focus on finite-dimensional inverse problems and interpret their in- and outputs as realizations of random variables.

3.2.1 Bayesian modeling

When considering inverse problems of the form (3.1), which involve finite-dimensional operators $\widehat{K} : \mathbb{R}^n \to \mathbb{R}^m$, it can be beneficial to construe their exact data $x^\Diamond$ and their exact solution $y^\Diamond$, i.e., the entities that fulfill

$$\widehat{K} y^\Diamond = x^\Diamond, \tag{3.6}$$

as realizations of the random variables X and Y (cf., e.g., [Stuart, 2010]). With respect to the data this interpretation can be intuitively justified by the emergence of noisy measurements. From this perspective it also becomes clear that $x^\Diamond$ should not only be seen as an arbitrary realization of X but beyond that should represent its expected value. The consideration of the random variable Y becomes meaningful as we take into account that the solution of real world inverse problems always lies in an anticipated range. When reconstructing the physical measurements recorded during the tomography of a brain, it is for example a lot more likely that the result of this inverse problem depicts some kind of general brain structure rather than a butterfly.

Thus, through the introduction of random variables we allow for a model margin which enables to incorporate additional information when solving the inverse problem. By choosing how to model the measurement noise and agreeing on a probability distribution of Y which models the available a-priori knowledge, we are able to tailor the solution approach of an inverse problem to its special needs.

To conceive how differently designed random variables affect this procedure explicitly, in the following we want to contemplate some examples. In doing so, we primarily want to focus on modeling different noise behaviors.

Gaussian noise

First of all we want to turn toward the case in which the noise introduced through inexact measurements is assumed to be additive and pointwise normally distributed with zero mean and variance $\sigma^2 > 0$. Together with our previous considerations this means that we can interpret the random variable X as the composition of m subordinate random variables X_i, $i \in \{1, \ldots, m\}$ with $X_i \sim \mathcal{N}(x_i^\Diamond, \sigma^2)$. Additionally presuming that these are pairwise independent and identically distributed, the probability of observing some arbitrary $x \in \mathbb{R}^m$ then amounts to

$$P(x) := P(X=x) = \prod_{i=1}^{m} P(X_i = x_i) = \prod_{i=1}^{m} \frac{\exp\left(-\frac{(x_i - x_i^\diamond)^2}{2\sigma^2}\right)}{\sqrt{2\pi}\sigma} = \frac{\exp\left(-\frac{\|x - x^\diamond\|_{\mathbb{R}^m}^2}{2\sigma^2}\right)}{\sqrt{2\pi}\sigma}.$$

Beyond that, for given exact solution $y^\diamond$ of (3.6) this probability can be specified to

$$P(x|y^\diamond) := P(X = x|Y = y^\diamond) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\|\widehat{K}y^\diamond - x\|_{\mathbb{R}^m}^2}{2\sigma^2}\right).$$

Having these terms at hand and aiming for an approximate solution of the corresponding inverse problem, an idea from statistics suggests to identify the *maximum likelihood* (ML) estimator, i.e., to find $\hat{y}$ such that observing x is most likely. In other words, this means to determine $\hat{y}$ such that

$$\hat{y} \in \arg\max_{y \in \mathbb{R}^n} P(x|y) = \arg\min_{y \in \mathbb{R}^n} -\log P(x|y) = \arg\min_{y \in \mathbb{R}^n} \frac{1}{2\sigma^2} \|\widehat{K}y - x\|_{\mathbb{R}^m}^2.$$

Unfortunately, this approach does not seem to be too promising as we regress to finding a least squares solution, which, recalling the observations in the previous section, could lead to severe problems.

However, looking from a different stochastic angle a better approach can be achieved. Instead of aiming for the ML estimator it is beneficial to engage with the search for the *maximum a-posteriori probability* (MAP) estimator. Here, trying to determine the element $\hat{y}$ which most likely generated the given data x, one addresses the maximization of the conditional probability $P(y|x)$. Respecting Bayes' rule, which states that

$$P(y|x) = \frac{P(x|y)P(x)}{P(y)},$$

this translates to identifying $\hat{y}$ such that

$$\begin{aligned} \hat{y} \in \arg\min_{y \in \mathbb{R}^n} -\log P(y|x) &= \arg\min_{y \in \mathbb{R}^n} -\log P(x|y) - \log P(y) \\ &= \arg\min_{y \in \mathbb{R}^n} \frac{1}{2\sigma^2} \|\widehat{K}y - x\|_{\mathbb{R}^m}^2 - \log P(y). \end{aligned}$$

Now assuming that the probability distribution of Y obeys a *Gibbs prior* [Geman and Geman, 1984], i.e., that

$$P(y) = c \exp(-\alpha \mathcal{R}(y))$$

for a normalizing constant $c > 0$ and some parameter $\alpha > 0$, this expression can be simplified to

$$\hat{y} \in \arg\min_{y \in \mathbb{R}^n} \frac{1}{2\sigma^2} \|\widehat{K}y - x\|_{\mathbb{R}^m}^2 + \alpha \mathcal{R}(y). \tag{3.7}$$

Note that this actually reminds of the formulation in (3.5) and even recovers the Tikhonov regularization (3.4) for $\mathcal{R}(y) = \|y\|_{\mathbb{R}^n}^2/2$. Hence, we have derived a minimization problem which yields a solution to (3.2) while adapting to the fact that the measured data includes additive Gaussian noise. Thereby we incorporated the accessible data-driven a-priori knowledge. Nevertheless, since we neglected modeling the random variable Y, i.e., we did not involve any additional information regarding the solution set, the regularizing functional $\mathcal{R}$ remains unspecified.

Poisson noise

Similar to the case of additive Gaussian noise, it is also possible to contemplate the scenario in which the exact data are corrupted by Poisson noise. To do so we need to restrict ourselves to inverse problems of the form (3.2) whose associated entity $x^\Diamond$ lies in $\mathbb{N}^m$. This for example is the case when dealing with positron emission tomography, as here during the measurement process one counts the number of photons in a section of a predefined grid (cf., e.g., [Wernick and Aarsvold, 2004]). Analogously to our previous considerations we then can interpret the random variable X as the composition of pairwise independent and identically distributed random variables X_i, $i = \{1, \dots m\}$, which now obey the following conditional distribution:

$$P(x_i | y^\Diamond) = \frac{(\widehat{K}y^\Diamond)_i^{x_i} \exp(-(\widehat{K}y^\Diamond)_i)}{x_i!}.$$

Note that with this choice we again made sure that the pointwise mean coincides with $x_i^\Diamond = (\widehat{K}y^\Diamond)_i$. Now following the same line of argument as before, in this scenario the MAP estimator $\hat{y}$ can be found by examining

$$\begin{aligned} \hat{y} \in \arg\min_{y \in \mathbb{R}^n} -\log P(y|x) &= \arg\min_{y \in \mathbb{R}^n} -\log \prod_{i=1}^m \frac{(\widehat{K}y)_i^{x_i} \exp(-(\widehat{K}y)_i)}{x_i!} - \log P(y) \\ &= \arg\min_{y \in \mathbb{R}^n} \sum_{i=1}^m (\widehat{K}y)_i - x_i \log(\widehat{K}y)_i + \alpha \mathcal{R}(y). \end{aligned}$$

This expression confirms, that modeling the measurement noise implies which data fidelity term $\mathcal{D}_x$ suits the considered inverse problem.

According to these two derivations also a lot of other occuring types of noise allow for the direct determination of suitable measures for the data discrepancy. To name some of these we want to mention Gamma noise, Laplace noise and Speckle noise. However, other types exist whose emergence is more difficult to deal with. For problems involving salt-and-pepper noise

for example we are not able to identify direct instructions to find their MAP estimator.
Hence, one of the two main tasks in variational modeling is to design adequate data fidelity terms for noise scenarios which elude the previous implementations. The second one is concerned with modeling the random variable Y, i.e., with finding ways to incorporate information about the solution set by defining appropriate regularizing functionals $\mathcal{R}$.

3.2.2 Transition to the infinite-dimensional case

In the previous subsection we had to restrict ourselves to the consideration of finite-dimensional problems in order to be able to interpret the entities x and y in (3.2) as realizations of random variables and derive explicit data fidelity terms. This raises the question of how these terms have to look like when dealing with infinite-dimensional problems.
In fact, there exist several heuristics which justify that transferring the discrete concepts to the continuous setting is a reasonable approach. The most simple one is probably to argue that the previous derivation of measures for the data discrepancy only motivates the definition of specified functionals $\mathcal{D}_x$. Since most devices measure fixed data and no random variable they just serve as an orientation. For a more detailed discussion about this transition we refer to [Dashti et al., 2013] and [Helin and Burger, 2015].
Hence, with this short remark we want to put on record that it is common practice to use the continuous counterparts to the respective derived data fidelity terms when contemplating infinite-dimensional inverse problems. In the scenario in which signal-independent additive Gaussian noise with zero mean and variation σ^2 is expected this for instance means to examine the minimization problem

$$\hat{y} \in \arg\min_{y \in \mathcal{Y}} \frac{1}{2} \|Ky - x\|_{\mathcal{X}}^2 + \alpha \mathcal{R}(y).$$

3.2.3 Example: Sparse signal reconstruction

After understanding how the design of minimization problems can support solving inverse problems, we now want to illustrate this process with a practical example. To do so we want to contemplate the scenario in which one is interested in recovering a sparse one-dimensional signal $y^\Diamond \in \mathbb{R}^n$.
For this purpose we presume that we are given the data $x \in \mathbb{R}^n$ which represent a noisy version of the convolution of the sought-after signal with a specified kernel. Hence, our interest lies in *deconvolving* x, or more precisely, in solving

$$\widehat{K} y^\Diamond = x, \tag{3.8}$$

with $\widehat{K} : \mathbb{R}^n \to \mathbb{R}^n$ representing the discrete convolution operator. Under the assumption that the corruption of the measured data was provoked by additive Gaussian noise with zero mean and variance $\sigma^2 = 1$, we are convinced by our previous implementations that a minimization problem corresponding to (3.8) should involve

$$\mathcal{D}_x(\widehat{K}y) = \frac{1}{2}\|\widehat{K}y - x\|_{\mathbb{R}^n}^2$$

as a guarantee for data fidelity. Now to respect the a-priori information given on $y^\Diamond$, i.e., incorporating that most of its entries can be expected to equal zero, a first idea could be to additionally include the so-called ℓ^0-norm,

$$\|y\|_0 := \sum_{j=1}^{n} |y_j|^0 \quad \text{with} \quad 0^0 := 0,$$

as a regularizing term. And in fact, by counting the number of non-zero entries its incorporation in the constrained problem

$$\min_{y\in\mathbb{R}^n} \|y\|_0 \qquad \text{s. t.} \qquad \widehat{K}y = x \tag{3.9}$$

is viable. However, involving this pseudo-norm in a minimization problem of the form (3.5), which then, due to the non-convexity of the ℓ^0-norm, could only be solved in a combinatorial fashion, poses difficulties. In particular, it was shown that problems of this kind are NP-hard to solve (see, e.g., [Fornasier, 2010]).
In order to overcome this issue, in the 1990s Tibshirani instead proposed to regularize with the convex relaxation of the ℓ^0-norm, namely the ℓ^1-norm (cf. [Tibshirani, 1996]). And indeed, being the largest convex function below the ℓ^0-norm, its application, among all ℓ^p-norms with $p \geq 1$, penalizes deviations from the trivial signal most. Beyond that, there also exist several less heuristic reasons justifying this approach. Especially, there was derived a variety of explicit sufficient conditions ensuring that a solution of (3.9) coincides with the unique solution to

$$\min_{y\in\mathbb{R}^n} \|y\|_1 \qquad \text{s. t.} \qquad \widehat{K}y = x.$$

To name only a few of them, we here want to mention the mutual incoherence property (cf. [Donoho and Huo, 2001]), the nullspace property (cf. [Cohen et al., 2009]), the exact recovery condition (cf. [Tropp, 2004]) and the restricted isometry property (cf. [Candès and Tao, 2005]).
Now, returning to our initial deconvolution problem we can conclude from the preceeding considerations, that in order to solve (3.8) it is a good idea to examine

$$\min_{y\in\mathbb{R}^n} \frac{1}{2}\|\widehat{K}y - x\|_{\mathbb{R}^n}^2 + \alpha\|y\|_1,$$

which is commonly called the LASSO method.
Note that, outside the application to toy examples, this method is particularly important in the field of compressed sensing (see [Donoho, 2006; Candès et al., 2006a,b]). Here, one classically wants to solve a superordinate inverse problem whose solution is assumed to be a

linear combination of a few basis elements. With access to this dictionary it therefore only remains to find the coefficients generating the solution. Hence, by transforming the originally involved operator to work on these parameters itself, a modified inverse problem can be defined. Due to the presumed sparsity regarding the vector of coefficients, this is then predestinated for the application of the LASSO method. Via this workaround it is thus possible to reduce the degrees of freedom corresponding to the original problem and find reasonable solutions although reverting to a smaller amount of data samples.

To conclude the present example, we furthermore want to point out that, at first glance, its translation to the infinite-dimensional setting turns out to be much more difficult. Since in the naturally chosen space of functions, the space $L^1(\Omega)$, elements with single non-zero values represent a null set, the regularization with $\| \,.\, \|_{L^1(\Omega)}$ does not seem to be very promising. In fact, the modeling of such problems requires the transition to the space of finite Radon measures (see, e.g., [Bredies and Pikkarainen, 2013; Boyer et al., 2017]).

3.2.4 Existence of solutions

Within the course of this chapter we have already learned that defining suitable energy functionals $\mathcal{F} : \mathcal{Y} \to \overline{\mathbb{R}}$ and examining their associated minimization problems of the form (3.5) can be a good idea to circumvent the difficulties that come with ill-posed inverse problems. Nevertheless, up to now we neglected discussing what exactly is meant when speaking about 'suitable' functionals.

In the light of the aforementioned it stands to reason that this suitability refers to the modeling aspect, i.e., it denominates the demand for data fidelity and regularizing terms which model the considered scenario sufficiently well. However, beyond that a very crucial criterion should be that the selected functional has a minimizer after all.

In the following we want to focus on this latter, more tangible aspect and find properties ensuring its validity. While doing so we define more closely which type of minimization problem is worth to contemplate when aiming for an approximate solution to problems of the form (3.2).

Thinking of the existence of minimizers of functionals $\mathcal{F}$ mapping to $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ a very first concern relates to the extensional characteristic. If there does not exist any argument $y \in \mathcal{Y}$ which generates a function value different from infinity, no minimizer can be determined. Therefore, in order to eliminate this risk we want to restrict ourselves to considering functions whose range involves at least one element in $\mathbb{R}$.

Definition 3.4. (cf., e.g., [Ekeland and Temam, 1999, Def. 1.4])
Let $\mathcal{F} : \mathcal{Y} \to \overline{\mathbb{R}}$ be a functional which attains a finite value for at least one argument, i.e., there exists $y \in \mathcal{Y}$ such that $\mathcal{F}(y) < \infty$. Then F is called *proper*.

Additional to the claim of being proper it seems reasonable to prevent that $\mathcal{F}$ attains its minimum at the 'borders' of its domain, i.e., for values $y \in \mathcal{Y}$ with large associated norm. In order to formalize this idea we want to introduce the concept of *coercivity*.

Definition 3.5. (cf., e.g., [Bauschke and Combettes, 2011, Def. 11.11])
A function $\mathcal{F} : \mathcal{Y} \to \overline{\mathbb{R}}$ is said to be *coercive*, if $F(y) \to +\infty$ as soon as $\|y\|_{\mathcal{Y}} \to +\infty$.

Similar to this property the above-mentioned constraint can also be expressed through the boundedness of the corresponding sublevel sets,

$$\operatorname{lev}_{\leq \xi} \mathcal{F} := \{y \in \mathcal{Y} | \mathcal{F}(y) \leq \xi\}.$$

In fact, even the following proposition applies.

Proposition 3.6. (cf., e.g., [Bauschke and Combettes, 2011, Prop. 11.12])
A functional $\mathcal{F} : \mathcal{Y} \to \overline{\mathbb{R}}$ is coercive if and only if its sublevel set $\operatorname{lev}_{\leq \xi} \mathcal{F}$ is bounded for any $\xi \in \mathbb{R}$.

Thus, through the introduction of proper and coercive functionals we can already guarantee, that the considered $\mathcal{F}$ has bounded sublevel sets of which at least one is non-empty. Nevertheless, we still allow the functional to implicitly define minimizing sequences whose limits are not taken. Since this behavior clearly complicates the determination of a minimizer, we want to suppress its occurrence by defining the subsequent final characteristic.

Definition 3.7. (cf., e.g., [Ekeland and Temam, 1999, Def. 1.21])
A functional $\mathcal{F} : \mathcal{Y} \to \overline{\mathbb{R}}$ is said to be (sequentially) *lower semi-continuous in* $y \in \mathcal{Y}$ if

$$\mathcal{F}(y) \leq \liminf_{\gamma \to \infty} \mathcal{F}(y_\gamma)$$

applies for all convergent sequences $y_\gamma \to y$. If this property holds for every $y \in \mathcal{Y}$, $\mathcal{F}$ is called (sequentially) *lower semi-continuous.*

Indeed, additional to our visual intuition, it can also be shown rigorously that the claim for lower semicontinuity resolves the issue with non-closed sublevel sets.

Lemma 3.8. (cf., e.g., [Bauschke and Combettes, 2011, Lem. 1.36])
For $\mathcal{F} : \mathcal{Y} \to \overline{\mathbb{R}}$ the following two statements are equivalent:

(i) $\mathcal{F}$ is lower semicontinuous.

(ii) For all $\xi \in \mathbb{R}$, the sublevel set $\operatorname{lev}_{\leq \xi} \mathcal{F}$ is closed in $\mathcal{Y}$.

With this last assertion at hand we finally have an idea of how a guideline ensuring the existence of minimizers of $\mathcal{F}$ could look like. However, we notice that this guideline is not completly clear yet, since within Definition 3.7 we did not specify the topology with respect to which the mentioned convergence is to be understood. This leaves room for interpretation and, in particular, needs us to choose a suitable setting on our own.
While thinking about an appropriate choice, we quickly notice that in general the strong (norm) topology on $\mathcal{Y}$ does not meet our requirements. This is because in infinite dimensions, a

direct consequence of Riesz's Lemma [Rudin, 1991, Lem. 4.22] states that here the closed unit ball is not compact. Hence, it is difficult to find converging sequences. Unfortunately, also with respect to the weak topology on $\mathcal{Y}$ we encounter problems ensuring this general compactness. Nevertheless, a remedy can be found by considering the following result by Banach and Alaoglu.

Theorem 3.9. (*Banach-Alaoglu*) (cf., e.g., [Megginson, 2012])
Suppose that $\mathcal{Z}$ is a Banach space and let $\mathcal{Y} = \mathcal{Z}^*$ be its dual space. Then, for any constant $c > 0$ the ball

$$B_c \coloneqq \{y \in \mathcal{Y} \mid \|y\|_{\mathcal{Y}} \leq c\} \subset \mathcal{Y}$$

is compact with respect to the weak* topology.

The statement in this theorem clearly suggests to concentrate on the weak* topology when it comes to sequential lower semicontinuity. Beyond that, for reflexive Banach spaces $\mathcal{Y}$, i.e., if $\mathcal{Y}^{**} \cong \mathcal{Y}$ and the weak* and weak topology coincide, even the weak topology can be identified suitable.
With this insight we are now able to formulate the subsequent concluding existence theorem.

Theorem 3.10. Suppose that $\mathcal{Y} = \mathcal{Z}^*$ for a Banach space $\mathcal{Z}$. Furthermore, let $\mathcal{F} : \mathcal{Y} \to \overline{\mathbb{R}}$ be a proper, weakly* lower semicontinuous and coercive functional. Then, $\mathcal{F}$ attains a minimum in $\mathcal{Y}$.

Proof. Suppose that $\mathcal{F}$ is proper. Then, there exists $\tilde{y} \in \mathcal{Y}$ with $\mathcal{F}(\tilde{y}) < \infty$ and the sublevel set $\operatorname{lev}_{\leq \mathcal{F}(\tilde{y})} \mathcal{F}$ is well-defined and non-empty. Beyond that, together with the coercivity and weak* lower semicontinuity of $\mathcal{F}$, Proposition 3.6 and Lemma 3.8 ensure that this sublevel set is bounded and closed. Now let $(y_\phi)_{\phi \in \mathbb{N}} \subset \operatorname{lev}_{\leq \mathcal{F}(\tilde{y})} \mathcal{F}$ be a minimizing sequence, i.e.,

$$\lim_{\phi \to \infty} \mathcal{F}(y_\phi) = \inf_{y \in \operatorname{lev}_{\leq \mathcal{F}(\tilde{y})} \mathcal{F}} \mathcal{F}(y).$$

Then, according to the boundedness of $\operatorname{lev}_{\leq \mathcal{F}(\tilde{y})} \mathcal{F}$ and Banach-Alaoglu there exists a weakly* convergent subsequence $(y_{\phi_\psi})_{\psi \in \mathbb{N}} \subseteq (y_\phi)_{\phi \in \mathbb{N}}$ whose limit y_* lies in $\operatorname{lev}_{\leq \mathcal{F}(\tilde{y})} \mathcal{F}$ due to its weak* sequential closedness. Involving the weak* lower semicontinuity of $\mathcal{F}$ further implies, that

$$\inf_{y \in \operatorname{lev}_{\mathcal{F}(\tilde{y})} \mathcal{F}} \mathcal{F}(y) = \lim_{\psi \to \infty} \mathcal{F}(y_{\phi_\psi}) = \liminf_{\psi \to \infty} \mathcal{F}(y_{\phi_\psi}) \geq \mathcal{F}(y_*).$$

Hence, incorporating that $\mathcal{F}$ is proper, we can conclude that

$$\mathcal{F}(y_*) = \inf_{y \in \mathcal{Y}} F(y) =: F_* > -\infty$$

applies, which proves the assertion. ☐

Having this result at hand it now remains to verify the listed properties on an individual basis, in order to be sure that a particular, designed minimization problem can be considered qualified to find an approximate solution to the inverse problem of interest. The overall suitability of a customized minimization problem, however, still also depends on the less quantifiable modeling aspect.

4

Variational methods for dynamic MRI

In the previous chapter we briefly introduced the concept of inverse problems and their connection to variational methods. In the following, we now want to discuss how we can explicitly utilize these methods in the context of dynamic MRI. In doing so, we primarily present techniques for the reconstruction of undersampled MR scans that exploit the high temporal correlation of dynamic measurements.
To this end, we start the present chapter by highlighting the physical process of measuring an MR signal. After understanding which kind of output is to be expected when dealing with MRI, we then give a quick introduction into the specific (ill-posed) inverse problem which underlies the reconstruction of such measurements. Subsequently, we devote special attention to the definition of various matrix norms that can be employed to variational regularization and debate their respective benefits with regard to dynamic MRI. While doing so we also establish a new class of regularizers, namely the generalized nuclear norm. In a final step we then want to formulate application-oriented variational problems for the reconstruction of undersampled dynamic MR data that resort to these matrix norms and discuss the existence of their solutions.

4.1 Dynamic MRI: Physics and methodical limits

4.1.1 The measurement process

When speaking about magnetic resonance imaging (MRI) in a mathematical context we are not able to avoid mentioning the physical process its measurements are based on. Having to deal with the reconstruction of its raw data makes us first want to gain a deeper understanding of how they are acquired. Although a detailed explanation of the fairly complicated measurement process is beyond the scope of this thesis, we want to give a short simplified introduction to the main procedure. For more elaborated and complete in-depth information we refer to [Elster and Burdette, 2001] or [Liang and Lauterbur, 2000].

The signal, which is measured in MRI, is mainly based on hydrogen atomic nuclei and their so-called *nuclear spin*. Around its nuclei this spin creates a randomly oriented weak magnetic field that is equipped with an associated magnetic moment. Through the influence of an external strong magnetic field B_0 these magnetic moments start to lapse into a movement which is commonly referred to as *precession*: they randomly wobble around an axis which is aligned

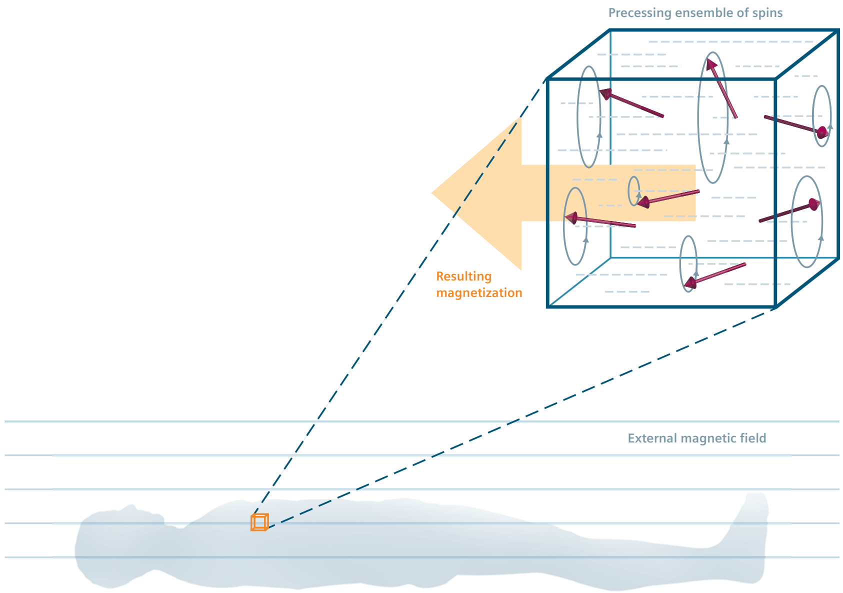


Figure 4.1: *Illustration of a voxels spin ensemble during the exposition to the external magnetic field B_0 (image courtesy of [Siemens Healthcare GmbH, 2015]).*

with the outer magnetization. Since the global magnetization of a collection of nuclei can be thought of as the vector sum of all individual magnetic moments, this alignment of all precession axes enables the detection of a macroscopic magnetic field. While the microscopic transversal magnetizations still cancel out due to the incoherent precession of the magnetic moments, the longitudinal magnetizations which are oriented parallel to the external strong magnetization add up to a weak signal. This situation describes the equilibrium state which serves as the base of operations with respect to MR measurements. A visualization can be found in Figure 4.1.

In order to generate an MR signal, in this state high-frequency radio pulse waves are emitted perpendicularly to the magnetic field B_0. Approximating the precessing frequency, the so-called *Larmor frequency*, these waves stimulate the magnetic moments to lapse into some kind of excited state. In this excited state the precession of all magnetic moments proceeds in phase and, as depicted in Figure 4.2, with an altered angle with respect to the direction of the external magnetic field B_0. Consequently, the previous longitudinally oriented global magnetization

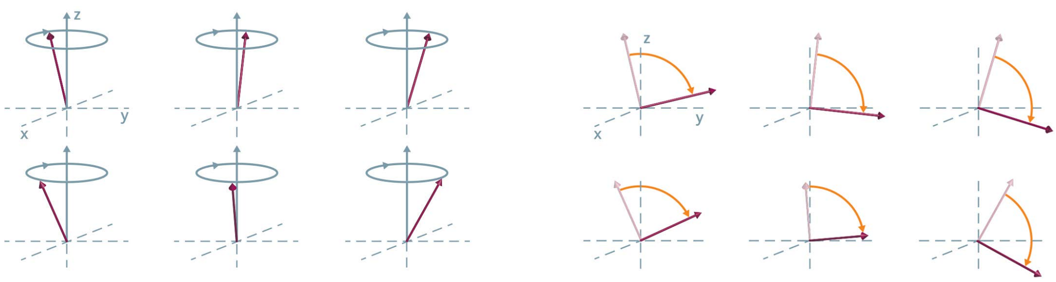


Figure 4.2: *Illustration of a spin ensemble before (left) and at the end of (right) a 90-degree pulse (image courtesy of [Siemens Healthcare GmbH, 2015]).*

transforms to a precessing transversal one which, due to the generated alternating electricity, can be measured.
As soon as the emission of the radio frequency waves is suspended, the magnetic moments gradually return to their lower energy state. During this relaxation now two simultaneously but independently happening processes can be observed. On the one hand the longitudinal magnetization is regaining strength, since the magnetic fields created by the nuclear spins again start to align with B_0. On the other hand the magnetic moments increasingly precess in an incoherent manner, which causes the transversal magnetization to shrink. The former observation is commonly called *T1 relaxation*, while the latter bears the name *T2 relaxation*. Following this procedure, measurements which are rich in contrast and thus eventually allow to differentiate between diverse tissues arise because the respective relaxation velocity primarily depends on the composition and structure of the tissue the nuclei belongs to.
By dint of spatial encoding techniques which, because of their complexity, are not discussed in detail, these measurements can be assigned to special locations in the frequency domain. This domain is commonly called *k-space*. Thus, at the end of an MR scan there is always a raw complex-valued data matrix whose entries are layer by layer based on the measurements gathered from the associated slice of the scanned object. However, it should be noted that there does not exist an one-to-one relation between these single entries and the voxels of the corresponding slice. Instead, every single entry contains partial information on every voxel of the considered layer.
In order to transfer this data from the frequency domain into an image that is interpretable for the human eye the *Fourier transform* proved to be a useful tool. Interpreting the entries in the k-space as encodings of frequencies which all together contribute to the visual representation of the physiological origin suggests to compute the discrete inverse Fourier transform of every layer.
Indeed, in the context of 'perfect' measurements this procedure leads to decent reconstructions. Though, when dealing with real medical devices we unfortunately can not always act on this assumption. In practice we rather have to expect data that is incomplete and/or corrupted by the random Brownian motion of molecules and inaccurate measurements.
In order to, in spite of these corruptions, obtain reconstructions which deviate only in a limited scope from the hypothetical reconstructions of the exact data, we want to stabilize their determination with the help of prior information. Recalling the techniques introduced in the previous chapter it stands to reason to incorporate these within the framework of variational methods.

4.1.2 The (ill-posed) inverse problem to dynamic MRI

On the basis of the described measurement process and recalling the concepts introduced in Chapter 3 it is obvious that the reconstruction of the data acquired during an MR scan amounts to solving an inverse problem. Moreover, we are convinced that in its most simple and general version the corresponding underlying operator equation should read

$$\widehat{K}y = x \tag{4.1}$$

with $\widehat{K} : \mathbb{C}^m \to \mathbb{C}^k$ characterizing the discrete Fourier transform defined by

$$(\widehat{K}y)_k := \frac{1}{m}\sum_{j=1}^{m} y_j \exp\left(-2\pi i \frac{(j-1)(k-1)}{m}\right). \tag{4.2}$$

Accordingly, $x \in \mathbb{C}^k$ can be interpreted as the noisy k-space measurements and $y \in \mathbb{C}^m$ represents the sought-after vectorized image consisting of m pixels.
Note that here, espescially in the case in which the sampling of the k-space is based on a Cartesian grid, the most nearby practice is to choose $m = k$. With that, the number of pixels to be reconstructed equals the number of Fourier space measurements and a well-defined linear system is created.
However, due to the very time-consuming and therefore expensive process of data acquisition, the reconstruction of MR measurements also qualifies for the application of the concept of compressed sensing (see, e.g., [Lustig et al., 2007]). As already insinuated in Section 3.2.3, here the objective is to find solutions to highly under-determined linear systems. Applied to the problem at hand, this means that reverting to significantly less k-space coefficients one still aims for the computation of images with high resolution, i.e., it rather holds that $m \gg k$. In this setting, in order to prevent measurements which are redundant with respect to the linear system and therefore even enhance the under-determined characteristic, it is a lot more important to guarantee their pairwise incoherence. Thus, one here usually resorts to sampling schemes other than the Cartesian one. Examples of this include the technique of radial sampling introduced by Lauterbur (see [Lauterbur, 1973]) and the method of spiral sampling (see, e.g., [Noll et al., 1995]). Nevertheless, even when dealing with these different sampling schemes typically the acquired data is retrospectively assigned to positions in an equidistant grid (see, e.g., [O'Sullivan, 1985; Jackson et al., 1991]) and describing the reconstruction process through (4.1) with (4.2) remains reasonable.
Especially in this latter setting we realize that, in order to reconstruct the data acquired during an MR scan, we may have to face an ill-posed inverse problem whose (approximate) solving needs the incorporation of additional information. Respecting the fact that in connection with MRI deviations introduced during the measurement process are commonly modeled as additive Gaussian noise with zero mean (see, e.g., [Elster and Burdette, 2001]), a suitable minimization problem for this task should be the following:

$$\arg\min_{y \in \mathbb{C}^m} \frac{1}{2}\|\widehat{K}y - x\|_{\mathbb{C}^k}^2 + \alpha\mathcal{R}(y).$$

Here $\alpha > 0$ represents the already introduced regularizing parameter and $\mathcal{R}$ a yet to be defined functional.

Within the further scope of this thesis we, however, predominantly want to concentrate on the problem we are confronted with when dealing with dynamic MRI. In contrast to the just described static MRI scenario, we here consider the task of simultaneously reconstructing a whole series of images. If we again interpret the single images as vectorized entities, which now are column by column combined in a matrix, the associated inverse problem can be described through

$$\widetilde{K}A = B. \tag{4.3}$$

In this formulation $A \in \mathbb{C}^{m\times n}$ represents the time series of n images, which each consist of m pixels, and $B \in \mathbb{C}^{k\times n}$ can be understood as the concatenation of the k-space measurements acquired at the n different points in time. Beyond that, the involved operator $\widetilde{K} : \mathbb{C}^{m\times n} \to \mathbb{C}^{k\times n}$ constitutes a composition of single subordinate operators, which each act on the single columns of the argument matrix as already defined in (4.2), i.e., for A_j denoting the jth column of matrix $A \in \mathbb{C}^{m\times n}$ it holds that

$$\widetilde{K}A := \begin{pmatrix} \widehat{K}A_1 & \dots & \widehat{K}A_n \end{pmatrix} \in \mathbb{C}^{k\times n}$$

(cf. [Xiang and Henkelman, 1993]).

Coming back to the issue of time-consuming measurements we have to realize that these hamper the acquisition of dynamic data even more than it was already the case in the static scenario. Having to deal with its aftereffects in this setting means to decide between a good spatial or temporal resolution. If we prioritize the latter, the spatial resolution suffers, because only a few measurements per frame remain to perform their reconstruction. The other way around, when aiming for a good spatial resolution we have to consult more data for the reconstruction of the individual frames, which evidently weakens the temporal resolution. Now, in order to nevertheless obtain high-resolution results with respect to both, time and space, in this dynamic setting it is particularly advisable to engage with the concept of compressed sensing. This again comprises to consider $m \gg k$ and include additional a-priori knowledge in order to compensate the lack of spatial information.

Therefore, in analogy to the static case and under the assumption of additive Gaussian noise with zero mean, it seems to be a good idea to realize the reconstruction in this dynamic setting by examining the following minimization problem:

$$\arg\min_{A\in\mathbb{C}^{m\times n}} \frac{1}{2}||\widetilde{K}A - B||_F^2 + \alpha\mathcal{R}(A). \tag{4.4}$$

Here $\|\,.\,\|_F$ indicates the *Frobenius norm* which, characterized via

$$\|A\|_F := \left(\sum_{i,j} |A_{ij}|^2 \right)^{\frac{1}{2}},$$

represents a matrix analogue to the Euclidean norm on vectors.
With this approach, in contrast to analyzing n individual static minimization problems, we then have access to a wider range of regularizing functionals $\mathcal{R}$. These can now not only refer to properties of the individual images but also relate to the whole series. Thereby, it is possible to exploit the clearly existing relation between the single frames and stabilize their reconstruction by incorporating this information into the minimization process. In order to explicitly find such suitable regularizing functionals $\mathcal{R}$, and therefore specify the general variational approach to the reconstruction of dynamic MR scans in (4.4), we, however, have to become aware of more concrete attributes assigned to the series of frames to be reconstructed.

4.2 Matrix norms in the context of dynamic MRI

In the previous section we took a physical view on an MR scan and found that, especially in the context of dynamic MRI, its very time-consuming characteristic represents a major drawback. Nevertheless, we also discovered that there is hope for remedy: Transforming the associated inverse problem to an ill-posed one while countervailing with the incorporation of additional information, promises to find satisfying approximate reconstructions.
To pursue the previously mentioned approach and find appropriate functionals $\mathcal{R}$ completing the variational method in (4.4), in this section we first of all want to introduce some norms, which operate on matrices and will play a relevant role when incorporating a-priori knowledge on a series of images. While doing so, we especially want to concentrate on the definition of a new class of norms, namely the generalized nuclear norm.

4.2.1 Mixed norms

To start this section, let us turn toward a class of matrix norms that we have already encountered in a special form while applying the Frobenius norm. We already noted, that this matrix norm can be understood as an analogue to the Euclidean norm, i.e., the ℓ^2-norm, on vectors. Now, in order to universalize this approach, we would like to transfer the concept of general ℓ^p-norms on vectors to matrices. For this purpose one can interpret the rows and columns of a matrix as vectors and, on them, perform a composition of their respective ℓ^p- and ℓ^q-norm.

Definition 4.1. (cf., e.g., [Benedek and Panzone, 1961; Samarah et al., 2005])
Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. Suppose that $\omega \in \mathbb{R}^{m \times n}$ with $\omega_{ij} > 0$ for all $i \in \{1, \ldots, m\}$, $j \in \{1, \ldots, n\}$

and $p, q \geq 1$. Then, the weighted *mixed* p, q*-norm* of $A \in \mathbb{K}^{m\times n}$ is defined via

$$\|A\|_{\omega;p,q} := \left(\sum_{i=1}^{m} \left(\sum_{j=1}^{n} \omega_{ij} |A_{ij}|^q \right)^{\frac{p}{q}} \right)^{\frac{1}{p}}.$$

If $\omega = \mathbb{1}$, with $\mathbb{1} \in \mathbb{R}^{m\times n}$ representing the matrix of ones, the weight-specifying notation can be dropped, i.e., $\|\,.\,\|_{p,q} := \|\,.\,\|_{\mathbb{1};p,q}$.

Note that here, the validity of all three norm criteria directly follows from the respective properties of the underlying vector norms. Besides, we realize that for $p = q = 2$ and $\omega = \mathbb{1}$ this mixed norm, in fact, coincides with the already implemented Frobenius norm. Beyond that, we perceive that, in the special case where $p = q = 1$ and again $\omega = \mathbb{1}$, applying the p, q-norm equals the compution of the ℓ^1-norm of the vectorized version of the matrix argument. And indeed, in the further course of this chapter we will see, that in the context of variational methods this particular mixed norm has to be treated in a similar way as the ℓ^1-norm.

4.2.2 Nuclear norm

Another possibility to deduce a matrix norm from the established vector norms, is to contemplate the singular value decomposition (SVD) of a matrix $A = U\Sigma V^* \in \mathbb{K}^{m\times n}$ into a unitary matrix $U \in \mathbb{K}^{m\times m}$, a diagonal matrix $\Sigma \in \mathbb{R}^{m\times n}$ and the (conjugate) transpose of a unitary matrix $V \in \mathbb{K}^{n\times n}$. Focusing on the resulting vector $\sigma \in \mathbb{R}^{\min(m,n)}$ of non-negative and unique diagonal entries of Σ, the singularvalues, then allows to identify a norm on A with a common ℓ^p-norm on σ. This results in the definition of the class of *Schatten p-norms* on matrices (cf., e.g., [Schatten, 2013]).

In the following we will see that, for our purposes, especially the application of the ℓ^1-norm on the singular values of a matrix proves useful. Thus, we here want to highlight the subsequent characterization.

Definition 4.2. (cf., e.g., [Schatten, 2013])
Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. Suppose that $\omega \in \mathbb{R}^{m\times n}$ with $\omega_{ij} > 0$ for all $i \in \{1, \ldots, m\}$, $j \in \{1, \ldots, n\}$ and $A \in \mathbb{K}^{m\times n}$. Let $(\sigma_\xi(\omega \cdot A))_{\xi\in\{1,\ldots,\min(m,n)\}}$ denote the singular values corresponding to the pointwise product of ω and A. Then, the weighted *nuclear norm* of A is given by

$$\|A\|_{\omega;*} := \sum_{\xi=1}^{\min(m,n)} \sigma_\xi(\omega \cdot A).$$

If $\omega = \mathbb{1}$, with $\mathbb{1} \in \mathbb{R}^{m\times n}$ representing the matrix of ones, the weight-specifying notation can be dropped, i.e., $\|\,.\,\|_* := \|\,.\,\|_{\mathbb{1};*}$.

Note that, respecting that the singular values of a matrix A coincide with the square roots of the eigenvalues corresponding to A^*A, respectively AA^*, we can verify that the above-mentioned expression indeed represents a norm.

Lemma 4.3. $\|\,.\,\|_{\omega;*}$ characterized as in Definition 4.2 fulfills all norm criteria.

Proof.

(i) *Positive definiteness:* Let $A \in \mathbb{K}^{m\times n}$ with $\|A\|_{\omega;*} = 0$. Then all singular values of $\omega \cdot A$ equal zero, which implies that $\omega \cdot A = 0$. Since $\omega_{ij} > 0$ for all $i \in \{1,\dots,m\}$, $j \in \{1,\dots,n\}$ we therefore get that $A = 0$.

(ii) *Absolute homogeneity:* Let $\sigma_{\xi'} \geq 0$ be an arbitrary singular value to $\omega \cdot A \in \mathbb{K}^{m\times n}$ with corresponding right-singular vector $v_{\xi'}$, i.e.,

$$(\omega \cdot A)^*(\omega \cdot A)v_{\xi'} = \sigma_{\xi'}^2 v_{\xi'}.$$

Then, for $\alpha \in \mathbb{K}$

$$(\omega \cdot \alpha A)^*(\omega \cdot \alpha A)v_{\xi'} = |\alpha|^2(\omega \cdot A)^*(\omega \cdot A)v_{\xi'} = |\alpha|^2\sigma_{\xi'}^2 v_{\xi'}$$

applies and $|\alpha|\sigma_{\xi'}$ can be identified as a singular value to $\omega \cdot \alpha A$. Hence,

$$\|\alpha A\|_{\omega;*} = \sum_{\xi=1}^{\min(m,n)} |\alpha|\sigma_\xi(\omega \cdot A) = |\alpha|\,\|A\|_{\omega;*}.$$

(iii) *Subadditivity:* In the 1950s Ky Fan proved that for matrices $A, B \in \mathbb{K}^{m\times n}$

$$\sum_{\xi=1}^{k} \sigma_\xi(A+B) \leq \sum_{\xi=1}^{k} \sigma_\xi(A) + \sum_{\xi=1}^{k} \sigma_\xi(B)$$

applies as soon as $k \leq \min(m,n)$ (cf. [Fan, 1951]). This directly implies the subadditivity of the nuclear norm.

□

In addition it may be noteworthy that, in Definition 4.2, we can determine even more precisely how many relevant singular values of A, respectively $\omega \cdot A$, have to be summed up in order to receive its nuclear norm. Considering that for arbitrary matrices $A \in \mathbb{K}^{m\times n}$, $B \in \mathbb{K}^{n\times k}$, $C \in \mathbb{K}^{l\times m}$ with $\operatorname{rank} B = n$ and $\operatorname{rank} C = m$ it holds that $\operatorname{rank} AB = \operatorname{rank} A = \operatorname{rank} CA$, we can infer from the unitarity of matrices U and V, which in both cases implies a full rank, that $\operatorname{rank} A = \operatorname{rank} \Sigma$. Since the latter, because of the diagonal shape of Σ, obviously coincides with the number of non-zero singular values, we can conclude, that

$$\|A\|_* = \sum_{\xi=1}^{\operatorname{rank} A} \sigma_\xi(A).$$

In fact, this connection between the number of non-zero singular values of a matrix and its rank will be of particular interest in the following section. Remembering that the rank of a matrix represents its number of linearly independent columns, it points to the fact that also the singular values are related to this linear independence.

4.2.3 Generalized nuclear norm

For the definition of the previously characterized nuclear norm we were reverting to the singular values of a matrix $A \in \mathbb{K}^{m\times n}$ to assign a norm to it. While doing so, respecting that these unique singular values represent the square roots of the eigenvalues to the matrix A^*A, respectively AA^*, we have to realize that the assigned norm value highly depends on the explicit shape of A^*. When speaking of an ordinary SVD we usually identify this entity with the (conjugate) transpose to A, i.e.,

$$A^* = A^H = \bar{A}^T.$$

Here, the notation making use of the asterisk, which usually indicates adjoint operators, is not entirely coincidental. In fact, with this definition we implicitly consider the matrix A to represent a linear function between the Hilbert spaces $\mathbb{K}^n$ and $\mathbb{K}^m$, both equipped with the standard inner product, and use its adjoint operator A^* for the computation of singular values. This insight raises the question, if it may also be interesting to contemplate some kind of modified singular values, which result from a different assumption.
Thus, in order to define a modified nuclear norm, henceforth we want to assume that the considered matrix A represents a linear function between the spaces $\mathbb{K}^n$ and $\mathbb{K}^m$, which are both equipped with alternative inner products. To specify these, we state the following characterization.

Definition 4.4. Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. Suppose that $B \in \mathbb{K}^{m\times m}$ is a Hermitian (symmetric) and positive definite matrix. Then, we define $\langle\,.\,,\,.\,\rangle_B : \mathbb{K}^m \times \mathbb{K}^m \to \mathbb{K}$ via

$$\langle x, y\rangle_B := x^H B y.$$

Note that, because of the properties of B, the just defined sesquilinear (bilinear) form defines an inner product on $\mathbb{K}^m$. Beyond that, it also induces a norm which we want to denote by $\|\,.\,\|_B$. Hence, assuming that $C \in \mathbb{K}^{m\times m}$ and $D \in \mathbb{K}^{n\times n}$ both fulfill the requirements in Definition 4.4, we can interpret the matrix $A \in \mathbb{K}^{m\times n}$ as a linear function between the Hilbert spaces $(\mathbb{K}^n, \langle\,.\,,\,.\,\rangle_D)$ and $(\mathbb{K}^m, \langle\,.\,,\,.\,\rangle_C)$. Under this assumption, the (conjugate) transpose to A does not meet the conditions of an adjoint operator anymore. Rather, it now applies for all $x \in \mathbb{K}^m$ and $y \in \mathbb{K}^n$ that

$$y^H A^H C x = \overline{x^H C A y} = \overline{\langle x, Ay\rangle_C} = \overline{\langle A^* x, y\rangle_D} = \overline{x^H (A^*)^H D y} = y^H D A^* x,$$

where the adjoint operator to A is again represented through a matrix. Together with the positive definiteness of D we can therefore conclude, that

$$A^* = D^{-1} A^H C.$$

As a consequence from the consideration of alternative Hilbert spaces we can, thus, obviously determine other eigenvalues and eigenvectors of A^*A, respectively AA^*. And indeed, it is even possible to define a little bit more precisely how these can be computed. Contemplating the new corresponding eigenproblem

$$A^* A v_\xi^{C,D} = \lambda_\xi^{C,D} v_\xi^{C,D} \quad \Leftrightarrow \quad D^{-1} A^H C A v_\xi^{C,D} = \lambda_\xi^{C,D} v_\xi^{C,D} \quad \Leftrightarrow \quad A^H C A v_\xi^{C,D} = \lambda_\xi^{C,D} D v_\xi^{C,D},$$

we can observe that, due to the hermiticity and positive definiteness of C and D, by means of substitution with $\tilde{v}_\xi = D^{\frac{1}{2}} v_\xi^{C,D}$

$$\left(C^{\frac{1}{2}} A D^{-\frac{1}{2}}\right)^H \left(C^{\frac{1}{2}} A D^{-\frac{1}{2}}\right) \tilde{v}_\xi = D^{-\frac{1}{2}} A^H C A D^{-\frac{1}{2}} \tilde{v}_\xi = \lambda_\xi^{C,D} \tilde{v}_\xi \tag{4.5}$$

holds. Consequently, the generalized singular values of A coincide exactly with the classical ones of the matrix $C^{\frac{1}{2}} A D^{-\frac{1}{2}}$. In addition, through (4.5) we can infer that the right-singular vectors $(\tilde{v}_\xi)_{\xi \in \{1,\dots,\min(m,n)\}}$ of $C^{\frac{1}{2}} A D^{-\frac{1}{2}}$ characterize the generalized right-singular vectors of A via $v_\xi^{C,D} = D^{-\frac{1}{2}} \tilde{v}_\xi$. Note that this identity ensures, that these vectors are orthonormal to each other with respect to the inner product induced by D, i.e., they behave according to the newly introduced structure. The corresponding assertion addressing the generalized left-singular vectors can be verified following the same line of argument while contemplating the eigenproblem to AA^*. Hence, defined via $u_\xi^{C,D} = C^{-\frac{1}{2}} \tilde{u}_\xi$ these also preserve the general structure of an SVD and, identifying $U_{C,D}$ and $V_{C,D}$ as the matrices whose columns represent the respective singular vectors, in summary we can state that

$$\left(V_{C,D}\right)^H D V_{C,D} = I, \quad \left(U_{C,D}\right)^H C U_{C,D} = I.$$

However, we should notice that, since in general

$$V_{C,D} D \left(V_{C,D}\right)^H \neq I, \quad U_{C,D} C \left(U_{C,D}\right)^H \neq I$$

applies, in contrast to the classical SVD the rows of $U_{C,D}$ and $V_{C,D}$ can not assumed to be orthonormal regarding the inner product induced by C, respectively D. Nevertheless, due to the consistency of row and column rank, we can be sure that the rows of $U_{C,D}$ and $V_{C,D}$ are still linearly independent with respect to each other.
All in all the above construction results in the subsequent matrix decomposition.

Proposition 4.5. Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. Suppose that $C \in \mathbb{K}^{m \times m}$, $D \in \mathbb{K}^{n \times n}$ are Hermitian (symmetric) and positive definite matrices and that $A \in \mathbb{K}^{m \times n}$. Then, there is a diagonal matrix $\Sigma_{C,D} \in \mathbb{R}^{m\times n}$ with diagonal entries $\sigma_\xi^{C,D} \geq 0$ for $\xi \in \{1, \dots, \min(m,n)\}$ and there exist matrices $U_{C,D} \in \mathbb{K}^{m\times m}$, $V_{C,D} \in \mathbb{K}^{n\times n}$ with

$$(V_{C,D})^H DV_{C,D} = I, \quad (U_{C,D})^H CU_{C,D} = I,$$

such that $A = U_{C,D}\Sigma_{C,D}(V_{C,D})^H D$ applies. A decomposition of this type is called the by C and D induced *generalized singular value decomposition* and $(\sigma_\xi^{C,D})_{\xi\in\{1,\dots,\min(m,n)\}}$ are called the *generalized singular values* of A.

Proof. In accordance with the previous considerations, the existence of this decomposition directly follows from the existence of the classical SVD (see, e.g., [Horn and Johnson, 1994, Thrm. 3.1.1]) of the matrix $C^{\frac{1}{2}}AD^{-\frac{1}{2}}$. Deducing that

$$C^{\frac{1}{2}}AD^{-\frac{1}{2}} = \tilde{U}\Sigma_{C,D}\tilde{V}^H \quad \Leftrightarrow \quad A = \left(C^{-\frac{1}{2}}\tilde{U}\right)\Sigma_{C,D}\left(D^{-\frac{1}{2}}\tilde{V}\right)^H D$$

for unitary (orthogonal) matrices $\tilde{U} \in \mathbb{K}^{m\times m}$, $\tilde{V} \in \mathbb{K}^{n\times n}$ leads to canonical candidates which fulfill the required conditions. □

It may be noteworthy that, although we here had to correct the decomposition of A by the additional incorporation of the matrix D, with this construct we still recover the common matrix equation

$$AV_{C,D} = U_{C,D}\Sigma_{C,D}.$$

Now, coming back to our original idea, based on this generalized SVD we want to define a more generalized version of the nuclear norm.

Definition 4.6. Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. Suppose that $C \in \mathbb{K}^{m\times m}$, $D \in \mathbb{K}^{n\times n}$ are Hermitian (symmetric) and positive definite matrices. Furthermore, let $\omega \in \mathbb{R}^{m\times n}$ with $\omega_{ij} > 0$ for all $i \in \{1, \dots, m\}$, $j \in \{1, \dots, n\}$ and $A \in \mathbb{K}^{m\times n}$. Let $(\sigma_\xi^{C,D}(\omega \cdot A))_{\xi\in\{1,\dots,\min(m,n)\}}$ denote the by C and D induced generalized singular values of the pointwise product of ω and A. Then, the weighted *generalized nuclear norm* of A is defined through

$$\|A\|_{\omega;*_{C,D}} := \sum_{\xi=1}^{\min(m,n)} \sigma_\xi^{C,D}(\omega \cdot A).$$

If $\omega = \mathbb{1}$, with $\mathbb{1} \in \mathbb{R}^{m\times n}$ representing the matrix of ones, the weight-specifying notation can be dropped, i.e., $\|\,.\,\|_{*_{C,D}} := \|\,.\,\|_{\mathbb{1};*_{C,D}}$.

From Lemma 4.3 it follows by very straight-forward arguments that also in this case the denomination as a norm is justified.

Lemma 4.7. $\|\,.\,\|_{\omega;*_{C,D}}$ characterized as in Definition 4.6 fulfills all norm criteria.

Proof. While the exact same arguments as in the proof of Lemma 4.3 lead to the positive definiteness and absolute homogeneity of $\|\,.\,\|_{\omega;*_{C,D}}$, to prove its subadditivity we can exploit that the generalized singular values coincide with the classical ones of $C^{\frac{1}{2}}AD^{-\frac{1}{2}}$. Profiting from this relation, the subadditivity of the conventional nuclear norm can be transferred to its generalized version via

$$\begin{aligned}\|A+B\|_{\omega;*_{C,D}} &= \|C^{\frac{1}{2}}AD^{-\frac{1}{2}} + C^{\frac{1}{2}}BD^{-\frac{1}{2}}\|_{\omega;*} \\ &\leq \|C^{\frac{1}{2}}AD^{-\frac{1}{2}}\|_{\omega;*} + \|C^{\frac{1}{2}}BD^{-\frac{1}{2}}\|_{\omega;*} \\ &= \|A\|_{\omega;*_{C,D}} + \|B\|_{\omega;*_{C,D}}.\end{aligned}$$

□

By means of Definition 4.6 we hence defined a new class of norms, which, dependent on matrices C and D, opens up new possibilities while keeping a reliable connection to the well-studied standard nuclear norm. As we will see later on, this connection proves extremely useful, when dealing with the generalized nuclear norm in a numerical context.

4.3 Variational methods for the reconstruction of dynamics

With the norms defined in the previous section we now have all tools at hand to take a closer look at specifications of the minimization problem in (4.4). More precisely, we are now ready to contemplate properties of a series of tomographic images and, based on them, derive explicit regularizing functionals which support the reconstruction of undersampled dynamic MR data. To do so we especially want to concentrate on the models introduced in [Haldar and Liang, 2010] and [Otazo et al., 2015] and, through the incorporation of the generalized nuclear norm, find a generalization of the former.

4.3.1 Linear dependence among frames

In order to start establishing a program which is able to face the approximate solving of the ill-posed inverse problem in (4.3), we first of all want to turn toward the assumption of linearly dependent frames.

Thinking of the desired solution which should arise from a dynamic tomographic scan, we can state that, due to the common physical origin, this should consist of a series of images which share a lot of information and only differ with respect to a few innovations introduced in each of them. This observation can be translated to a high linear dependence among frames. Now, being aware of this characteristic, it stands to reason to incorporate this a-priori information into the reconstruction process and thereby exclude solutions that do not possess this property in order to reduce the existing degrees of freedom. Recalling the considerations in Section 4.2.2

we moreover know that, regarding the spatio-temporal matrix representing the sought-after series, this additional knowledge manifests in a small rank. Hence, thinking of the problem in (4.4), a canonical choice for a regularizing functional $\mathcal{R}$ would be the one which assigns its rank to a matrix and thus forces the minimization to favor solutions with low rank.
A problem of this form gained a lot of attention under the name the *Netflix problem* (cf. [ACM SIGKDD and Netflix, 2007]). Here, one wanted to face the matrix completion problem arising from fragmentarily filled film rating surveys, i.e., to supplement rating matrices whose columns were representing individual users and whose rows were constituting single films. Since one was assuming that the preference or taste of a user is only affected by very few factors and therefore can be described through the linear combination of only a few other opinions, also here one wanted to circumvent the ill-posedness of the associated inverse problem by favoring solutions with highly linear dependent columns, i.e., with low rank.
However, regularizing with the rank of a matrix implicitly corresponds to the regularization with the ℓ^0-norm of the singular values of a matrix. Unfortunately, in Section 3.2.3 we have already seen that, due to its non-convexity, the incorporation of this ℓ^0-norm causes difficulties. And in fact, also in this highly related case it can be shown, that using the rank-functional as a regularizer induces an NP-hard problem (cf. [Yue and So, 2016]).
Fortunately, similar to the scenario in which we demanded the sparsity of the argument itself, there is hope for remedy. Indeed, as before, the relaxation with the ℓ^1-norm is the key: Respecting that the nuclear norm, i.e., the ℓ^1-norm on the singular values, is the tightest convex relaxation of the matrix rank (cf. [Fazel et al., 2001]), it seems reasonable to identify an associated regularizer with this previously defined functional (cf., e.g., [Candès and Recht, 2009; Recht et al., 2010; Candès and Tao, 2010]). With that, we now explicitly favor the occurrence of as few non-zero singular values as possible.
Of course, coming back to our original problem, also with respect to dynamical reconstructions it makes sense to replace the generic regularizer in (4.4) with the nuclear norm, i.e., to consider

$$\arg \min_{A \in \mathbb{C}^{m \times n}} \frac{1}{2} ||\widetilde{K}A - B||_F^2 + \alpha \|A\|_*. \tag{4.6}$$

And in fact, from an interpretive point of view, in this setting the application of the nuclear norm appears to be even more reasonable: Reverting to the concepts of *principal component analysis* (PCA) (see, e.g., [Jolliffe, 2002]) we realize that through the SVD of a spatio-temporal matrix we are essentially identifying the principal dynamics arising in the corresponding series of images. This becomes particularly clear, when noting that for $A \in \mathbb{C}^{m \times n}$ with SVD

$$A = U \Sigma V^H$$

and $U = (u_1, \ldots, u_m)$ with $(u_i)_{i \in \{1,\ldots,m\}} \subset \mathbb{C}^m$, $V = (v_1, \ldots, v_n)$ with $(v_j)_{j \in \{1,\ldots,n\}} \subset \mathbb{C}^n$,

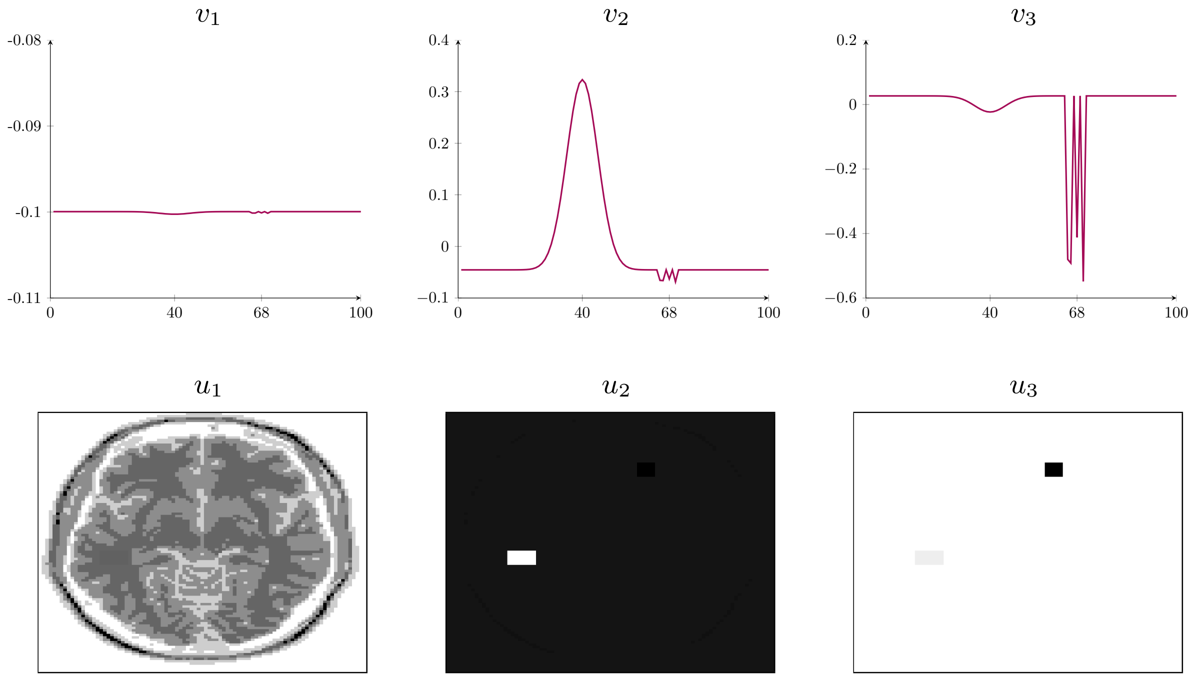


Figure 4.3: *Illustration of the first three left- and right-singular vectors of the spatio-temporal matrix corresponding to the time series of 100 steps created from stacking a phantom from the BrainWeb database [Cocosco et al., 1997] and introducing the following artificial dynamics: lower left corner – dynamic centered in time step* 40 *and evolving through a Gaussian kernel with standard deviation* $\lambda = 5$*; upper right corner – dynamic obeying a jump function for time steps* 65–70*. Associated singular values:* $\sigma_1 = 1$*,* $\sigma_2 = 0.0075$*,* $\sigma_3 = 0.0038$*.*

$\Sigma = \text{diag}(\sigma_1, \dots \sigma_{\min(m,n)})$ with $(\sigma_\xi)_{\xi \in \{1,\dots,\min(m,n)\}} \subset \mathbb{R}^+$ it holds that

$$A = \sum_{\xi=1}^{\min(m,n)} u_\xi \sigma_\xi v_\xi^H.$$

Here, while the vectors u_ξ localize the areas which are affected by a certain dynamic, the vectors v_ξ represent their temporal evolution. Simultaneously, their magnitude or variance gets specified through the corresponding singular value σ_ξ. Consequently, assuming that the singular values in Σ were organized in a descending order, i.e., that $\sigma_1 \geq \dots \geq \sigma_{\min(m,n)} \geq 0$, the SVD allows to represent a time series of images as the superposition of orthogonal dynamics assorted according to their significance and impact. An illustration of this can be found in Figure 4.3. Here, we can perfectly see, that for the present example the singular vectors corresponding to the largest singular value represent the more or less constant 'dynamic' of the background, which, compared to both other introduced dynamics, in fact describes the dominant characteristic of the time series. Besides, we observe that, while the singular vectors associated with the second largest singular value predominantly depict the smoothly generated dynamic in the lower left part of the series, the ones corresponding to the third singular value primarily characterize the non-smooth behavior in the upper right part. However, even though the number of non-zero

singular values for this example exactly coincides with the number of present dynamics, we also notice that the SVD does not allow to exactly separate them from each other.
Altogether, coming back to our assumption that the reconstructed image series of a dynamic tomographic scan should only include very few individual dynamics, it turns out to be even more advisable to penalize the presence of a lot of non-zero singular values. In fact, we will see in Chapter 5, that for the solution of problems of the form (4.6) the elimination of small singular values, i.e., of subordinate dynamics, will play a relevant role.

With this, we have now found various heuristic motivations for the incorporation of a regularizing nuclear norm. Now, to conclude this subsection we also briefly want to comment on their mathematical justification, i.e., on the existence of a solution to (4.6).

Lemma 4.8. Let $\mathcal{F}_\alpha : \mathbb{C}^{m\times n} \to \overline{\mathbb{R}}$ be the functional to be minimized in (4.6). Then, $\mathcal{F}_\alpha$ attains a minimum in $\mathbb{C}^{m\times n}$.

Proof. First of all it is obvious that $\mathcal{F}_\alpha$ is proper. Second, for a sequence $(A_\gamma)_{\gamma\in\mathbb{N}} \subset \mathbb{C}^{m\times n}$ whose Frobenius (or any other) norm converges toward infinity, we can be sure that due to the equivalence of all finite norms also $(\mathcal{F}_\alpha(A_\gamma))_{\gamma\in\mathbb{N}}$ reflects this behavior, ensuring the coercivity of $\mathcal{F}_\alpha$. And lastly, respecting that all norms are continuous and $\widetilde{K}$ is defined to be continuous, it is clear that as a concatenation of these elements also $\mathcal{F}_\alpha$ is continuous, i.e., in particular lower semicontinuous. With that, Theorem 3.10 guarantees the existence of a minimizer in $\mathbb{C}^{m\times n}$. □

4.3.2 Superposition of background and dynamics

In the previous subsection we introduced a regularizing technique which, based on the assumption that in the context of dynamic MRI the columns of a spatio-temporal matrix should be highly linearly dependent, involves the nuclear norm into the minimization problem in (4.4). Nevertheless, respecting the above-mentioned reflections on the superposition of dynamics, this approach could also bear a problem: Due to the claim for a too strong linear dependence, small but significant and worth to image dynamics could be eliminated during the reconstruction process.
A slightly different approach, therefore, suggests to exploit the presumed superposition of dynamics in a more explicit way (cf. [Chandrasekaran et al., 2011; Candès et al., 2011; Gao et al., 2012; Otazo et al., 2015]). If we are reminiscing about the singular values encoding the significance of the individual dynamics, we are convinced that the largest singular value σ_1 of a spatio-temporal matrix representing a tomographic series of frames should, more or less, be assigned to the background, i.e., to the physiological base. Beyond that, the associated right-singular vector v_1 should only consist of entries that have almost the same value.
With this consideration, it should also be possible to exploit the fact, that the sought-after reconstruction of the measured data can be separated into a background and some (true)

dynamics, i.e., to include the a-priori information

$$A = L + S.$$

Here, the matrix $L \in \mathbb{C}^{m\times n}$ representing the background should be characterized through a low rank, which can, but not necessarily needs to, equal one in order to also allow periodic dynamics to be counted as background. On the contrary, the matrix $S \in \mathbb{C}^{m\times n}$ describing the arising dynamics should, in comparison with L, contribute much less information to the entire series of images. This means that, with respect to a certain basis, this entity should be sparse. In the special case in which it is expected that the dynamics to be observed refer exclusively to a few voxels, it can even be presumed that S itself is of sparse shape.
Now, in order to explicitly involve this additional knowledge into the process of reconstruction, we should first of all realize that, strictly speaking, we are now confronted with a slightly modified inverse problem. Since the matrix A we are looking for is uniquely defined as soon as L and S are identified, the solution of the present inverse problem now reduces to the determination of these two components. According to this, we are now contemplating the operator equation

$$\check{K}\begin{pmatrix} L \\ S \end{pmatrix} = B,$$

where $\check{K} = \widetilde{K} \circ T$ with $T : \mathbb{C}^{m\times n} \times \mathbb{C}^{m\times n} \to \mathbb{C}^{m\times n}$ defined through $T(L, S) \coloneqq L + S$, i.e.,

$$\check{K} : \ \mathbb{C}^{m\times n} \times \mathbb{C}^{m\times n} \xrightarrow{T} \mathbb{C}^{m\times n} \xrightarrow{\widetilde{K}} \mathbb{C}^{k\times n}.$$

Thus, respecting the considerations in Sections 3.2.3 and 4.3.1 to incorporate the above-mentioned assumptions on low rank and sparsity, it seems reasonable to, in accordance with the minimization problem in (4.4), examine the following expression:

$$\arg \min_{L,S\in\mathbb{C}^{m\times n}} \frac{1}{2}\|\widetilde{K}(L + S) - B\|_F^2 + \beta_1\|L\|_* + \beta_2\|S\|_{1,1}. \tag{4.7}$$

Notice that, here, we are actually minimizing over the pair $(L, S) \in \mathbb{C}^{m\times n} \times \mathbb{C}^{m\times n}$, and the regularizer $\mathcal{R}$ mentioned in (4.4) is now understood to obey

$$\mathcal{R} : \mathbb{C}^{m\times n} \times \mathbb{C}^{m\times n} \to \overline{\mathbb{R}}, \quad \mathcal{R}(L, S) = \beta_1\|L\|_* + \beta_2\|S\|_{1,1}$$

for new parameters $\beta_1, \beta_2 > 0$. Simultaneously, the original regularizing parameter α is set to one.
Moreover, we should note that for the uniqueness, and therefore well-posedness, of a superpositional task like the one in (4.7) a further incoherence condition has to be satisfied. In addition to the one between the frequency domain (k-space) and the image domain, which is indispensable for the challenge of compressed sensing, we now also have to make sure to avoid identifiability issues. To do so, having a concrete application of a minimization problem of the form (4.7) in

mind we always have to ensure that the component which is identified via its low rank not itself has a sparse representation and vice versa. However, while in [Chandrasekaran et al., 2011] and [Candès et al., 2011] this additional incoherence constraint is elaborated on in more detail, by contemplating a specific scenario this requirement can often be considered fulfilled following a line of heuristic arguments.

Now, to conclude this subsection, as before, we want to guarantee that our intuitive derivation is also eligible in a mathematical sense. Hence, we ultimately want to address the existence of solutions to the minimization problem in (4.7).

Lemma 4.9. Let $\mathcal{F}_\alpha : \mathbb{C}^{m\times n} \times \mathbb{C}^{m\times n} \to \overline{\mathbb{R}}$ be the functional to be minimized in (4.7). Then, $\mathcal{F}_\alpha$ attains a minimum in $\mathbb{C}^{m\times n} \times \mathbb{C}^{m\times n}$.

Proof. Since the concatenation of two continuous functions preserves the continuity and a norm of an element in $\mathbb{C}^{m\times n} \times \mathbb{C}^{m\times n}$ tends to infinity if and only if the sum of any norms of its two components does so, we are convinced that, due to the same argumentation as in Lemma 4.8, $\mathcal{F}_\alpha$ has a minimum in $\mathbb{C}^{m\times n} \times \mathbb{C}^{m\times n}$. □

4.3.3 Linear dependence favoring smooth dynamics

The reconstruction techniques introduced in the precedent two subsections were both mainly based on the regularization with the standard nuclear norm. The motivation for this was the aspiration for a preferably high linear dependence among the frames of a time series, respectively for a preferably low rank of the associated spatio-temporal matrix. We thus used the nuclear norm as a convex relaxation of the matrix rank. However, remembering the previously introduced generalized nuclear norm which is based on the generalized SVD

$$A = U_{C,D}\Sigma_{C,D}\left(V_{C,D}\right)^H D,$$

we realize that the rank of a matrix A likewise coincides with the number of non-zero generalized singular values. This is essentially because the positive definiteness of matrices always implies a full rank. With this, having in mind that the matrices $U_{C,D}$ and $V_{C,D}$ can be represented through $U_{C,D} = C^{-\frac{1}{2}}\tilde{U}$ and $V_{C,D} = D^{-\frac{1}{2}}\tilde{V}$ with unitary $\tilde{U}$ and $\tilde{V}$ (see Section 4.2.3), it is clear that also $U_{C,D}$ and $V_{C,D}$ have full rank and the rank of A is exclusively characterized through the shape of the diagonal matrix $\Sigma_{C,D}$. This raises the question of whether, for suitable positive definite and Hermitian matrices C and D, it might also be reasonable to establish the generalized nuclear norm as a convex relaxation of the matrix rank.
In order to understand more precisely how this could affect the reconstruction process, we again want to engage with the decomposition of A in more detail. Similar to the classical SVD, also for the generalized version it applies that a matrix A can always be written as the superposition

of the outer products of its singular vectors, i.e., that

$$A = U_{C,D}\Sigma_{C,D}\left(V_{C,D}\right)^H D = \sum_{\xi=1}^{\min(m,n)} u_\xi^{C,D}\sigma_\xi^{C,D}\left(v_\xi^{C,D}\right)^H D. \tag{4.8}$$

So, even in this case, reverting to the concepts in PCA, we naturally decompose the given time series into its observable dynamics. These, again, get graded in their relevance through the magnitude of their associated singular values $\sigma_\xi^{C,D}$. And in fact, this graduation is exactly the essential point which, in the context of regularization, distinguishes the generalized SVD from the classical one. Considering the natural matrix norm, i.e., the one which interprets a matrix as an operator and subsequently assigns its operator norm, we agree that its explicit behavior significantly depends on the structures of the vector spaces which were assumed to underlie the associated image and preimage space. In particular, assuming that these two are the respective vector spaces induced by C and D results in

$$\|A\|^2 = \max_{\|y\|_D=1} \|Ay\|_C^2 = \max_{\|y\|_D=1} \langle y, A^*Ay\rangle_D.$$

Now, involving that with the C and D induced generalized SVD

$$A^*AV_{C,D} = V_{C,D}\Sigma_{C,D}^2 \quad \Leftrightarrow \quad (V_{C,D})^H DA^*AV_{C,D} = \Sigma_{C,D}^2$$

applies, together with the substitution $x = (V_{C,D})^H Dy$ we can infer, that

$$\begin{aligned}\|A\|^2 &= \max_{\|V_{C,D}x\|_D=1} \langle V_{C,D}x, A^*AV_{C,D}x\rangle_D \\ &= \max_{\|x\|_2=1} x^H (V_{C,D})^H DA^*AV_{C,D}x \\ &= \max_{\|x\|_2=1} \sum_{\xi=1}^{\min(m,n)} \left(\sigma_\xi^{C,D}\right)^2 |x_\xi|^2 \\ &= \left(\sigma_1^{C,D}\right)^2.\end{aligned}$$

Hence, by exploiting the identity

$$\|V_{C,D}x\|_D = x^H(V_{C,D})^H DV_{C,D}x = \|x\|_2$$

we realize, that, similar to the spectral norm, the natural matrix norm of A can be identified with its maximum singular value $\sigma_1^{C,D}$. This insight implies that

$$\max_{y\neq 0} \frac{\|Ay\|_C}{\|y\|_D} = \sigma_1^{C,D},$$

or in general that for every singular value $\sigma_\xi^{C,D}$ of A at least one vector $\bar{y} \in \mathbb{C}^n$ can be found satisfying

$$\frac{\|A\bar{y}\|_C}{\|\bar{y}\|_D} = \sigma_\xi^{C,D}. \tag{4.9}$$

Of course, above all, this is true for the in $(\mathbb{C}^n, \langle\,.\,,\,.\,\rangle_D)$ normed associated right-singular vectors $v_\xi^{C,D}$. However, beyond that the assertion in (4.9) is also valid for all non-normed versions of these vectors. And especially with respect to those it becomes clear, that the generalized singular values directly depend on the norms assigned to the image and preimage space.
To shed some light on this relation and to understand how this could possibly influence the reconstruction process, in the following we want to turn toward an explicit example: We want to commit ourselves to the scenario in which the image space of matrix A is equipped with the Euclidean norm while the preimage space coincides with the vector space which through Definition 4.4 is induced by an approximation of the negative Laplacian matrix. In concrete terms this means that from now on we are presuming that $C = I_m \in \mathbb{R}^{m\times m}$ and for $I_n, L \in \mathbb{R}^{n\times n}$, $\varepsilon > 0$ we are defining that

$$D = I_n - \varepsilon L \qquad \text{with} \quad L_{jj'} = \begin{cases} -2 & \text{if } j = j' \\ 1 & \text{if } |j - j'| = 1 \\ 0 & \text{else} \end{cases}. \tag{4.10}$$

Note that at this point the usage of the negative Laplacian matrix itself is not possible since it does not fulfill the required positive definiteness and thus turns out to be unsuitable to induce an inner product. Nevertheless, in order to use a matrix with very similar characteristics, we are able to consider the above-mentioned symmetric and positive definite approximation. Here, it is obvious that the bigger the parameter ε is chosen the better (but also 'less positive definite') this approximation gets.
Now, coming back to our previous considerations, with this example at hand the insight in (4.9) indeed becomes more tangible. With respect to the present setting we can now realize that the norm which appears in the numerator, as in the case of a classical SVD, coincides with the Euclidean one. However, the one which can be found in the denominator is characterized through the inner product that, with the help of Definition 4.4, is induced by the just characterized approximation of the negative Laplacian. According to that we can approximately identify it with a discretized H^1-norm. With that the expression in (4.9) suggests that the singular values, whose associated vectors $\bar{y} \in \mathbb{C}^n$ have a big discretized H^1-norm, should be rather small. Regarding the application in dynamic MRI this for example can be the case as soon as the singular value of interest is associated with a very non-smooth dynamic. In contrast to this, smoother dynamics which own a smaller discretized H^1-norm should be assigned to comparably bigger singular values. Beyond that, respecting that for the classical SVD the denominator in (4.9) includes the Euclidean norm which does not distinguish that much between smooth and non-smooth signals, with the help of the present generalized SVD it should be possible

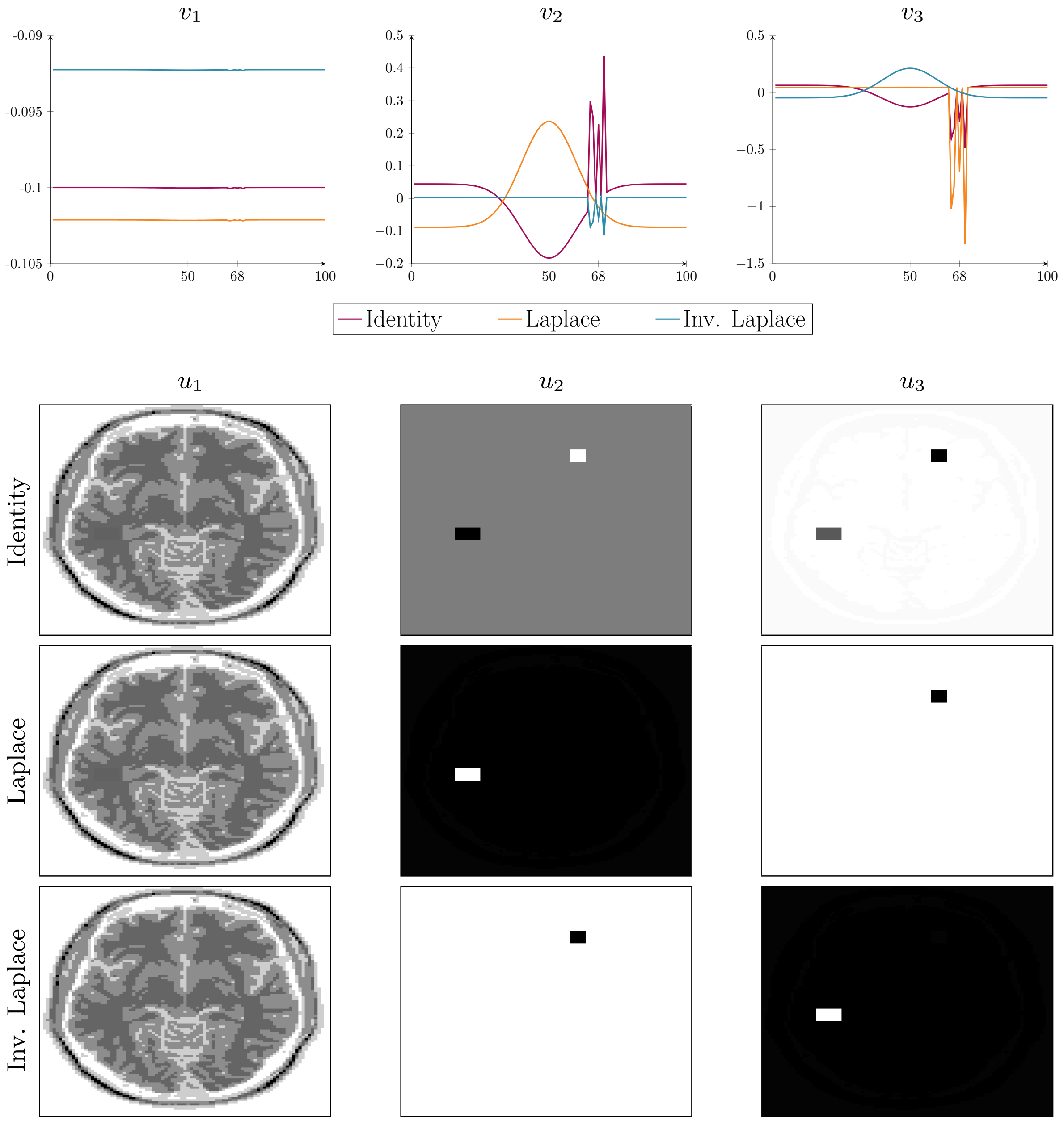


Figure 4.4: *Illustration of the first three left- and right-singular vectors corresponding to the generalized SVD with $C = I_m$ and $D \in \mathbb{R}^{n \times n}$ equaling the identity I_n minus the approximate negative Laplacian matrix (see (4.10) with $\varepsilon = 20$), or the inverse of the approximate negative Laplacian matrix, respectively. Time series underlying the considered spatio-temporal matrix features two dynamics: lower left corner – dynamic centered in time step 50 and evolving through a Gaussian kernel with standard deviation $\lambda = 10$; upper right corner – dynamic obeying a jump function for time steps 65–70. Associated singular values can be found in Table 4.1.*

to separate these kind of signals more clearly from each other. And indeed, in Figure 4.4 we can observe exactly this behavior. Here, for the singular vectors which emerge from the classical SVD, i.e., for which in the given example $D = I_n$ was chosen, due to the temporal

	σ_1	σ_2	σ_3
Identity	1	0.0024	0.0023
Laplace	0.9592	0.0021	0.0011
Inv. Laplace	1.1831	0.0135	0.0027

Table 4.1: *Listing of singular values corresponding to the experiment in Figure 4.4.*

overlapping of both imposed dynamics, it is hard to individually identify them. In contrast to that, the generalized SVD for which D approximates the negative Laplacian matrix succeeds well in this. However, note that, owing to the decomposition in (4.8), in this latter scenario we do not contemplate the generalized right-singular vectors $v_1^{C,D}$, $v_2^{C,D}$, $v_3^{C,D}$ themselves but their A-reproducing versions $Dv_1^{C,D}$, $Dv_2^{C,D}$, $Dv_3^{C,D}$. With them we are then able to observe the already presumed favoritism of smooth dynamics: Associated with the second singular value $\sigma_2^{C,D}$ the continuously constructed dynamic is perceived to be more dominant than the non-continuous one. In a logical constant way a contrary effect can be noticed when considering the generalized SVD induced by the inverse of the approximate negative Laplacian matrix. In this setting, big corresponding norm values are assigned to smoother signals such that these are eventually associated with smaller singular values.

Having this behavior in mind we can conclude that regularizing with the generalized nuclear norm, which is based on an approximation of the negative Laplacian, appears to be especially reasonable as soon as the resulting reconstruction is expected to feature smooth dynamics. This is because, as already foreshadowed, in Chapter 5 we will see that solving problems of the form (4.4) involving the classical nuclear norm is decisively based on the elimination of subordinate dynamics. Hence, transfering this conception to the generalized nuclear norm case, a more dominant perception of smooth dynamics should result in the preferential removal of non-smooth signals. A reconstruction emerging from a minimization problem like this should thus behave rather continuously over time. Note that, following this train of thought, this approach could also reduce the occurrence of noise due to its usually non-continuous temporal representation. At the same time, this is also the reason why considering (4.4) including the generalized nuclear norm based on the inverse of the approximate Laplacian matrix does not seem to be very beneficial in the context of dynamic imaging.

All in all we can summarize that, when anticipating a reconstruction incorporating smooth dynamics, it could be advantageous, instead of regularizing with the standard nuclear norm, to include as a regularizer the generalized nuclear norm induced by $C = I_m$ and $D \in \mathbb{R}^{n\times n}$ obeying the characterization in (4.10), i.e., to examine

$$\arg \min_{A\in\mathbb{C}^{m\times n}} \frac{1}{2}\|\widetilde{K}A - B\|_F^2 + \alpha\|A\|_{*_{C,D}}. \tag{4.11}$$

With that, in addition to involving prior knowledge on the linear dependence among the frames of the time series, we hope to simultaneously impose a smooth temporal evolution of occurring dynamics. Notice that, equivalently to the reasoning in Lemma 4.8, also here the existence of a solution in $\mathbb{C}^{m\times n}$ is obviously given.

With the minimization problems introduced in this section we now have some promising approaches at hand allowing to solve the ill-posed inverse problem to undersampled dynamic MRI. In order to validate their effectiveness, in the following chapter we want to turn toward their explicit solving and visualize their influence on the reconstruction process by means of application-related examples.

5 Numerical implementation and results

In the previous chapter we were predominantly engaged with the modeling of three different minimization problems which, in spite of fragmentary data, are able to yield stable reconstructions of dynamic MR measurements.
In this chapter we now want to turn toward their numerical solution and test their effectiveness with respect to explicit reconstruction problems. To do so, we first of all take a quick look on a general approach which enables to find minimizers to (partially) non-differentiable but in some sense 'simple' functionals, namely the forward-backward splitting, and subsequently derive corresponding algorithms fitted to the optimization problems described above. Thereafter, we contemplate three different concrete examples for their application: We start by addressing the scenario in which one is interested in reconstructing a series of frames which should depict the motion of very small cells. Subsequently, we realize that the approach in Section 4.3.2 also qualifies to track these kind of cells in series already reconstructed. To conclude we focus on the reconstruction of dynamic MR data representing a few smooth dynamics. While doing so, we are confronted with (temporary) limitations which, however, enable us to uncover some uncertainties regarding the established regularization with the classical nuclear norm as soon as a smoothing operator is involved.

5.1 Algorithmic solution

Within the precedent chapter we mainly spoke about the modeling of individual minimization problems, but neglected how these can be solved in practice. A field that has turned to this issue and furthermore has developed a precise analysis on the topic is the one of *convex optimization*. Since covering the vast findings in this field is clearly beyond the scope of this thesis, for a detailed discussion on this subject we here only want to refer to the elaborations in [Chambolle and Pock, 2016] and [Rasch, 2018, Chap. 3]. Beyond that for insights into the related field of convex analysis we recommend consulting the work of [Bauschke and Combettes, 2011].
However, at this point it shall suffice to revert to a very shortened and heuristic argumentation in order to deduce a numerical method which is capable of finding solutions to the previously introduced problems.

A very first intuitive idea to determine the minimizer $y_* \in \mathcal{Y}$ of a proper, convex and lower semicontinuous functional $\mathcal{F} : \mathcal{Y} \to \overline{\mathbb{R}}$ is probably to fix a point $y_0 \in \mathcal{Y}$ and from there iteratively move toward y_* by following the direction of the negative gradient. Though, in order to implement this approach in practice, it is first of all necessary to transfer the concept of differentiability to the present type of functionals.

Definition 5.1. (cf., e.g., [Bauschke and Combettes, 2011, Def. 2.45])
Let $\mathcal{F} : \mathcal{Y} \to \overline{\mathbb{R}}$ be proper. Then, $\mathcal{F}$ is said to be *Fréchet differentiable* in $y \in \mathcal{Y}$ if there exists a functional $\nabla\mathcal{F}(y) \in L(\mathcal{Y}, \mathbb{R})$ such that

$$\lim_{\|y'\|_{\mathcal{Y}} \to 0} \frac{\mathcal{F}(y + y') - \mathcal{F}(y) - \langle \nabla\mathcal{F}(y), y' \rangle}{\|y'\|_{\mathcal{Y}}} = 0$$

applies.

Hence, considering a Fréchet differentiable functional $\mathcal{F}$, the above-mentioned intuitive idea can actually be formalized. For a given step width $\tau > 0$ its realization obeys the fixed-point iteration

$$y_{k+1} = y_k - \tau\nabla\mathcal{F}(y_k). \tag{5.1}$$

Firstly formulated in the 1840s by Cauchy, this method is called the *explicit gradient descent* algorithm (see [Cauchy, 1847]). In spite of its very simple and comprehensible motivation it is, however, also evident that it features limitations: As soon as the functional to be minimized is not differentiable, the method in (5.1) is not applicable.
A slightly different approach, which at first glance is as well based on the Fréchet differentiability of $\mathcal{F}$, but, due to its implicit formulation, promises to behave more stable, is the one of the *implicit gradient descent.* For a step width $\tau > 0$ its iterative procedure is given by

$$y_{k+1} = y_k - \tau\nabla\mathcal{F}(y_{k+1}).$$

This, in due consideration of

$$\frac{y_{k+1} - y_k}{\tau} + \nabla\mathcal{F}(y_{k+1}) = 0 \quad \Leftrightarrow \quad y_{k+1} = \arg\min_{y \in \mathcal{Y}} \frac{1}{2\tau}\|y - y_k\|^2 + \mathcal{F}(y),$$

implies, that here the iterative steps themselves can be understood as minimization problems with respect to a (now strictly) convex functional. In 1965 Moreau integrated these minimization problems into the characterization of an operator.

Definition 5.2. (cf. [Moreau, 1965])
Let $\mathcal{F} : \mathcal{Y} \to \overline{\mathbb{R}}$ be proper, convex and lower semicontinuous. Then, the *proximal operator* to $\mathcal{F}$ with stepsize $\tau > 0$ is defined via

$$\mathrm{prox}_{\tau\mathcal{F}} : \mathcal{Y} \to \mathcal{Y}, \quad y \mapsto \arg\min_{y' \in \mathcal{Y}} \frac{1}{2\tau}\|y' - y\|^2 + \mathcal{F}(y').$$

Note that the contemplation of this proximal operator is also reasonable for functionals which do not fulfill the requirements of Fréchet differentiability. In fact, there are some non-differentiable functionals, which are 'simple' enough that their proximal operator has a closed form solution. As we will see later on, examples for this are the ℓ^1-norm and the nuclear norm.
This last remark already gives a vague idea how to proceed in the case of non-differentiable functionals. And indeed, when dealing with problems that are composed of a differentiable and a non-differentiable part, i.e., when being interested in solving

$$\arg\min_{y \in \mathcal{Y}} \mathcal{F}(y) + \mathcal{E}(y) \tag{5.2}$$

with $\mathcal{F}$ Fréchet differentiable and $\mathcal{E}$ non-smooth, the proximal operator plays a relevant role. To see this, we first of all have to realize that here, due to the non-smoothness of $\mathcal{E}$, it is obviously not possible to apply the method of explicit or implicit gradient descent. Therefore, in this setting an alternative and weaker notion of differentiability has to be implemented.

Definition 5.3. (cf., e.g., [Bauschke and Combettes, 2011, Def. 16.1])
Let $\mathcal{E} : \mathcal{Y} \to \overline{\mathbb{R}}$ be proper and convex. Then,

$$\partial\mathcal{E}(y) := \{p \in \mathcal{Y} \,|\, \mathcal{E}(y') \geq \mathcal{E}(y) + \langle p, y' - y\rangle \text{ for all } y' \in \mathcal{Y}\}$$

defines the *subdifferential* of $\mathcal{E}$ at $y \in \mathcal{Y}$.

In fact, in analogy to differentiable functionals one can find, that the optimality of an element $y_* \in \mathcal{Y}$ with respect to a proper, convex and lower semicontinuous functional $\mathcal{E}$ is given if and only if the zero element lies in the associated subdifferential, i.e., it holds that

$$0 \in \partial\mathcal{E}(y_*) \quad \Leftrightarrow \quad y_* \in \arg\min_{y \in \mathcal{Y}} \mathcal{E}(y) \tag{5.3}$$

(cf., e.g., [Bauschke and Combettes, 2011, Thrm. 16.2]). Since for Fréchet differentiable functionals $\mathcal{F}$ it can be shown that $\partial\mathcal{F}(y) = \{\nabla\mathcal{F}\}$ (cf., e.g., [Bauschke and Combettes, 2011, Prop. 17.26]) this even coincides with the more established version of Fermat's rule. Incorporating that furthermore also $\partial(\mathcal{F} + \mathcal{E})(y) = \partial\mathcal{F}(y) + \partial\mathcal{E}(y)$ applies for all $y \in \mathrm{int}\,\mathrm{dom}\,\mathcal{F} \cap \mathrm{int}\,\mathrm{dom}\,\mathcal{E}$ (cf. [Bauschke and Combettes, 2011, Prop. 6.19, Thrm. 16.37]), for problems of the form (5.2) this means that optimality is reached in $y_* \in \mathcal{Y}$ as soon as

$$0 \in \nabla\mathcal{F}(y_*) + \partial\mathcal{E}(y_*).$$

Note that, due to the maximal monotonicity of $\partial\mathcal{E}$ and the concomitant well-definedness of the operator $(I + \tau\partial\mathcal{E})^{-1}$ (cf. [Bauschke and Combettes, 2011, Chap. 23]), for any $\tau > 0$ this formulation can be rewritten to

$$\left(\frac{1}{\tau}I + \partial\mathcal{E}\right)(y_*) \in \frac{1}{\tau}y_* - \nabla\mathcal{F}(y_*) \quad \Leftrightarrow \quad y_* \in (I + \tau\partial\mathcal{E})^{-1}\left(y_* - \tau\nabla\mathcal{F}(y_*)\right).$$

Simultaneously, when applying the optimality condition in (5.3) to the previously introduced proximal operator one can analogously deduce that

$$\operatorname{prox}_{\tau\mathcal{E}}(y) = y_* \quad \Leftrightarrow \quad y \in (I + \tau\mathcal{E})(y_*)$$

and we realize that solving the problem in (5.2) is equivalent to finding the fixed point $y_* \in \mathcal{Y}$ of the function

$$y \mapsto \operatorname{prox}_{\tau\mathcal{E}}\left(y - \tau\nabla\mathcal{F}(y)\right).$$

Contemplating the corresponding fixed-point iteration, these considerations lead to the algorithm of *forward-backward splitting*

$$y_{k+1} = \operatorname{prox}_{\tau\mathcal{E}}(y_k - \tau\nabla\mathcal{F}(y_k)) \tag{5.4}$$

[Lions and Mercier, 1979; Combettes and Wajs, 2005]. This now finally enables to concretely solve problems of the form (5.2) involving a nonsmooth but 'simple' operator $\mathcal{E}$.

Now, in order to adapt this algorithm to more explicit scenarios, it solely remains to determine the Fréchet derivative of the differentiable part and the proximal operator corresponding to the non-differentiable one. In the following we want to do this for the minimization problems introduced in Chapter 4.

Regularization with the nuclear norm

When contemplating the functional to be minimized in (4.6) we can immediately determine that its data fidelity part $\mathcal{D}_B : \mathbb{C}^{m\times n} \to \overline{\mathbb{R}}$ satisfies the requirements of Fréchet differentiability and, thus, can be assigned to the derivative

$$\nabla\mathcal{D}_B(A) = \widetilde{K}^*(\widetilde{K}A - B). \tag{5.5}$$

The regularizing part which includes the nuclear norm, however, evades this concept. Fortunately, as it was to be hoped after the above-mentioned introduction to the forward-backward splitting algorithm, the nuclear norm is characterized in a way simple enough to have an analytically identifiable associated proximal operator.

Lemma 5.4. (cf. [Cai et al., 2010, Thrm. 2.1])
Let $\tau > 0$ and $\mathcal{T}_\tau : \mathbb{C}^{m\times n} \to \mathbb{C}^{m\times n}$ be the soft-singular-value-thresholding operator which for an arbitrary matrix A with SVD $A = U\Sigma V^H$ and $\xi \in \{1, \dots, \min(m, n)\}$ is defined via

$$\mathcal{T}_\tau(A) := U \operatorname{diag}(\max(0, \sigma_\xi - \tau))\, V^H.$$

Then, it applies that

$$\mathcal{T}_\tau(A) = \arg \min_{A' \in \mathbb{C}^{m\times n}} \frac{1}{2}\|A' - A\|_F^2 + \tau\|A'\|_*. \tag{5.6}$$

Proof. In (5.3) we have noted, that for a proper and convex functional a sufficient criterion for being minimal in some point is the belonging of the zero element to the associated subdifferential. In the considered scenario this means that the minimum of $\mathcal{F}(A') = \frac{1}{2}\|A' - A\|_F^2 + \tau\|A'\|_*$ is attained in $\hat{A} \in \mathbb{C}^{m\times n}$ if and only if

$$0 \in \hat{A} - A + \tau\partial\|\hat{A}\|_*, \tag{5.7}$$

where through the findings in [Watson, 1992] and [Lewis, 2003] it is known that

$$\partial\|\hat{A}\|_* = \{UV^H + W \mid \hat{A} = U\Sigma V^H,\, W \in \mathbb{C}^{m\times n},\, U^H W = 0,\, WV = 0,\, \sigma_{\max}(W) \le 1\}. \tag{5.8}$$

On the other hand, we can observe that for a decomposition of the SVD of A in

$$A = U_1\Sigma_1V_1^H + U_2\Sigma_2V_2^H,$$

where U_1, V_1 (respectively U_2, V_2) represent the singular vectors whose corresponding singular values are bigger than (respectively smaller than or equal to) τ, it holds that

$$\mathcal{T}_\tau(A) = U_1(\Sigma_1 - \tau I)V_1^H$$

and therefore

$$A - \mathcal{T}_\tau(A) = \tau(U_1V_1^H + \tau^{-1}U_2\Sigma_2V_2^H).$$

Since in this formulation the maximum singular value of $W := \tau^{-1}U_2\Sigma_2V_2^H$ is now obviously smaller than or equal to 1 and in a natural way also $WV = 0$ and $U^HW = 0$ are fulfilled, we can deduce together with (5.8) that $A - \mathcal{T}_\tau(A) \in \tau\partial\|\mathcal{T}_\tau(A)\|_*$. Hence, with (5.7) we can identify $\mathcal{T}_\tau(A)$ as a minimizer of $\mathcal{F}$ and through the strict convexity of $\mathcal{F}$ confirm its uniqueness. □

With this we realize that applying the proximal operator associated with a scaled nuclear norm to a matrix $A \in \mathbb{C}^{m\times n}$ essentially boils down to the shrinkage or elimination of its singular values. Thinking of the interpretation of the SVD of spatio-temporal matrices mentioned in Chapter 4, also from an applied point of view this procedure has its justification: In this scenario,

while aiming for a low rank solution, i.e., a solution with only a few non-zero singular values, through the application of the soft-singular-value-thresholding operator we get rid of the less important, subordinate dynamics and concentrate only on the remaining few influential ones. In doing so, the scaling parameter $\tau > 0$ controls up to which level dynamics are interpreted to be subordinate.
Combining this result with the derivative computed in (5.5), the forward-backward splitting algorithm in (5.4) eventually suggests to solve the minimization problem in (4.6) by defining a suitable step size $\tau > 0$ and following the fairly simple iterative scheme

$$\begin{cases} \bar{A}_{k+1} = A_k - \tau \widetilde{K}^*(\widetilde{K}A_k - B), \\ A_{k+1} = \mathcal{T}_{\tau\alpha}(\bar{A}_{k+1}) \end{cases} \tag{5.9}$$

until convergence is reached (cf. [Cai et al., 2010]).

Partial regularization with the nuclear and the $1,1$-norm

Turning toward the minimization problem in (4.7) we realize that its structure is very similar to the problem just discussed. Reverting to the same kind of data fidelity measure, this parts Fréchet derivative can be computed accordingly while respecting the twofold dependence on the argument $(L, S) \in \mathbb{C}^{m\times n} \times \mathbb{C}^{m\times n}$, i.e., here

$$\nabla \mathcal{D}_B(L,S) = \Big(\widetilde{K}^*(\widetilde{K}(L+S) - B)\,,\, \widetilde{K}^*(\widetilde{K}(L+S) - B)\Big)\,.$$

Concerning the non-differentiable regularizing part $\mathcal{R}(L,S) = \beta_1\|L\|_* + \beta_2\|S\|_{1,1}$ we again want to make use of the proximal operator. With respect to the present setting we can actually see that this falls into two minimizing components which separately operate on both arguments, i.e.,

$$\operatorname{prox}_{\tau\mathcal{R}}(L,S) = \begin{pmatrix} \arg\min_{L'\in\mathbb{C}^{m\times n}} \frac{1}{2}\|L' - L\|_F^2 + \tau\beta_1\|L'\|_* \\ \arg\min_{S'\in\mathbb{C}^{m\times n}} \frac{1}{2}\|S' - S\|_F^2 + \tau\beta_2\|S'\|_{1,1} \end{pmatrix}.$$

Regarding the first component we already know from the previous subsection how its closed form solution has to look like. For the second one, however, we first of all only realize its coincidence with the proximal operator of the scaled $1,1$-norm. Favorably, also for this a representation specifying its operation can be found analytically.

Lemma 5.5. Let $\tau > 0$ and $\mathcal{S}_\tau : \mathbb{C}^{m\times n} \to \mathbb{C}^{m\times n}$ be the shrinkage-thresholding operator that for an arbitrary matrix A is characterized through

$$\mathcal{S}_\tau(A)_{ij} := \max(0, |A_{ij}| - \tau)\ \operatorname{sgn}(A_{ij}).$$

Then, it holds that

$$\mathcal{S}_\tau(A) = \arg\min_{A' \in \mathbb{C}^{m\times n}} \frac{1}{2}\|A' - A\|_F^2 + \tau\|A'\|_{1,1}.$$

Proof. Aiming for a minimizer $\hat{A} \in \mathbb{C}^{m\times n}$ of $\mathcal{F}(A') = \frac{1}{2}\|A' - A\|_F^2 + \tau\|A'\|_{1,1}$ we first of all notice that its determination can be done pointwise, meaning that we can compute its components $\hat{A}_{ij}$ separately. To do so we again follow (5.3) and consider the optimality condition

$$0 \in \hat{A}_{ij} - A_{ij} + \tau\partial|\hat{A}_{ij}|.$$

By incorporating that respecting the definition of the subdifferential we can ensure that

$$p \in \partial|\hat{A}_{ij}| \quad \Leftrightarrow \quad \begin{cases} p = \operatorname{sgn}(\hat{A}_{ij}), & \text{if } \hat{A}_{ij} \neq 0, \\ |p| \leq 1, & \text{else}, \end{cases}$$

this leaves room for three different cases: To begin with, $\hat{A}_{ij} > 0$ can represent a minimizer as soon as $0 = \hat{A}_{ij} - A_{ij} + \tau$ is satisfied, i.e., if $\hat{A}_{ij} = A_{ij} - \tau > 0$. On the other hand, $\hat{A}_{ij} < 0$ fulfills the optimality condition as $0 = \hat{A}_{ij} - A_{ij} - \tau$, i.e., if $\hat{A}_{ij} = A_{ij} + \tau < 0$. And lastly, for $\hat{A}_{ij} = 0$ the optimality condition can be met if $0 \in \hat{A}_{ij} - A_{ij} + [-\tau, \tau]$, i.e., as $|A_{ij}| \leq \tau$. Thus, combining these scenarios we can summarize that pointwise optimality is reached in $\hat{A}_{ij}$ if

$$\hat{A}_{ij} = \begin{cases} A_{ij} - \tau\operatorname{sgn}(A_{ij}), & |A_{ij}| > \tau, \\ 0, & |A_{ij}| \leq \tau. \end{cases}$$

Transferring this result to the entire matrix then proves the assertion. □

With this we are now again able to bring together all of our findings to assemble the forward-backward splitting algorithm tailored to the present minimization problem. In doing so, we realize that, here, the iterative scheme in (5.4) can be split into separate updates regarding the two involved components L and S: While with respect to the latter we mainly follow the iterative shrinkage-thresholding algorithm (ISTA) (cf., e.g., [Chambolle et al., 1998; Daubechies et al., 2004; Figueiredo and Nowak, 2003]), concerning the former we more or less abide by the iterative method introduced in the precedent subsection. All in all, for an appropriate choice of $\tau > 0$ this results in iterating:

$$\begin{cases} \bar{L}_{k+1} = L_k - \tau\widetilde{K}^*(\widetilde{K}(L_k + S_k) - B), \\ L_{k+1} = \mathcal{T}_{\tau\beta_1}(\bar{L}_{k+1}), \\ \bar{S}_{k+1} = S_k - \tau\widetilde{K}^*(\widetilde{K}(L_k + S_k) - B), \\ S_{k+1} = \mathcal{S}_{\tau\beta_2}(\bar{S}_{k+1}). \end{cases} \tag{5.10}$$

Regularization with the generalized nuclear norm

Focusing on the problem in (4.11) which involves the generalized nuclear norm as a regularizer, we expect to be able to derive an iterative minimizing scheme by following similar steps as in the scenario of the classical nuclear norm which we already considered. Hence, we first of all contemplate the more simple (denoising) problem

$$\arg \min_{A' \in \mathbb{C}^{m\times n}} \frac{1}{2}\|A' - A\|_F^2 + \tau \|A'\|_{*C,D}. \tag{5.11}$$

With regard to this, it then would be desirable to mimic the proof of Lemma 5.4 for a soft-singular-value-thresholding operator, which now resorts to the *generalized* SVD. However, before doing so, we can turn toward the slightly differently defined minimization problem

$$\arg \min_{A' \in \mathbb{C}^{m\times n}} \frac{1}{2}\|C^{\frac{1}{2}}(A' - A)D^{-\frac{1}{2}}\|_F^2 + \tau \|A'\|_{*C,D} \tag{5.12}$$

and note that in consideration of (4.5) an equivalent formulation reads

$$\arg \min_{A' \in \mathbb{C}^{m\times n}} \frac{1}{2}\|C^{\frac{1}{2}}(A' - A)D^{-\frac{1}{2}}\|_F^2 + \tau \|C^{\frac{1}{2}}A'D^{-\frac{1}{2}}\|_{*}. \tag{5.13}$$

Moreover, by means of substitution with $\bar{A} = C^{\frac{1}{2}}A'D^{-\frac{1}{2}}$ this can further be rewritten to

$$\arg \min_{\bar{A} \in \mathbb{C}^{m\times n}} \frac{1}{2}\|\bar{A} - C^{\frac{1}{2}}AD^{-\frac{1}{2}}\|_F^2 + \tau \|\bar{A}\|_{*}, \tag{5.14}$$

which strongly reminds of the problem examined in (5.6). And in fact, here, together with the necessary resubstitution, the associated statement in Lemma 5.4 allows to identify the minimizer A_* of (5.12) with

$$A_* = C^{-\frac{1}{2}}\mathcal{T}_\tau(C^{\frac{1}{2}}AD^{-\frac{1}{2}})D^{\frac{1}{2}}.$$

Now, again involving our considerations in (4.5) this means that

$$A_* = U_{C,D}\,\mathrm{diag}(\max(0, \sigma_\xi^{C,D} - \tau))\,(V_{C,D})^H D \qquad \text{with } A = U_{C,D}\Sigma_{C,D}(V_{C,D})^H D,$$

i.e., that finding a minimizer to the problem in (5.12) coincides with applying the generalized soft-singular-value-thresholding operator induced by C and D to the input matrix A. With this insight we have to realize that, contrary to our expectations, the functional in (5.11) can not be minimized by the same argument. However, from an interpretative point of view it still is to be expected, that its exact minimizer behaves in a very similar way. Besides, unlike for the preceding scenarios, in the present one we are fortunately not obliged to find an exact solution to (5.11) in order to derive a numerical scheme for solving (4.11). This is because once again exploiting the characteristic in (4.5) together with the above-mentioned substitution allows us to study the equivalent formulation

$$\arg \min_{\bar{A} \in \mathbb{C}^{m\times n}} \frac{1}{2} \|\widetilde{K}\left(C^{-\frac{1}{2}} \bar{A} D^{\frac{1}{2}}\right) - B\|_F^2 + \alpha \|\bar{A}\|_*$$

instead. The contemplation of this problem now makes it possible to, more or less, adopt the iterative method in (5.9). By adjusting the gradient descent step with respect to the adapted Fréchet derivative of the data fidelity term

$$\nabla \mathcal{D}_B(\bar{A}) = C^{-\frac{1}{2}} \widetilde{K}^* \left(\widetilde{K}\left(C^{-\frac{1}{2}} \bar{A} D^{\frac{1}{2}}\right) - B\right) D^{\frac{1}{2}},$$

an equivalent reasoning leads to the scheme

$$\begin{cases} \bar{A}'_{k+1} = \bar{A}_k - \tau C^{-\frac{1}{2}} \widetilde{K}^* \left(\widetilde{K}\left(C^{-\frac{1}{2}} \bar{A}_k D^{\frac{1}{2}}\right) - B\right) D^{\frac{1}{2}}, \\ \bar{A}_{k+1} = \mathcal{T}_{\tau\alpha}(\bar{A}'_{k+1}). \end{cases} \tag{5.15}$$

Of course, in order to obtain a solution of the original problem in (4.11), after the convergent performance of this algorithm we have to make sure to resubstitute the found minimizer $\bar{A}_*$ via $A_* = C^{-\frac{1}{2}} \bar{A}_* D^{\frac{1}{2}}$.

5.2 Computational experiments

After the derivation of numerical methods which allow to solve the optimization problems introduced in Chapter 4, we are now ready to eventually verify their effectiveness. To do so we apply the above-mentioned algorithms to concrete examples of dynamic MR data and especially concentrate on the context of neuroimmune cell imaging. However, note that the presented regularization approaches can also support reconstructions in numerous other examples even including applications outside the context of dynamic MRI.

5.2.1 Undersampling imposing a superposition of dynamics

To get started, we first of all want to devote ourselves to the reconstruction of undersampled dynamic MR data which should portray the motion of very small cells in the brain. This becomes relevant, for instance, when trying to examine the behavior of neuroimmune cells. Already at an early stage these can draw attention to neuroinflammations in the brain. Therefore, their observation can serve as a very valuable diagnostic tool for diseases like multiple sclerosis and Alzheimer's. Beyond that, the tracking of these cells can be employed to test the effectiveness of existing therapeutic approaches (see, e.g., [Hemmer et al., 2015; Sánchez et al., 2012; Masthoff et al., 2019]).

In order to simulate the raw data generated by an MR scan which was performed to identify these kind of cells, we here once again want to take a slice of the simulated MR phantom from the BrainWeb database (see [Cocosco et al., 1997]) as a starting point. Through the 20-fold

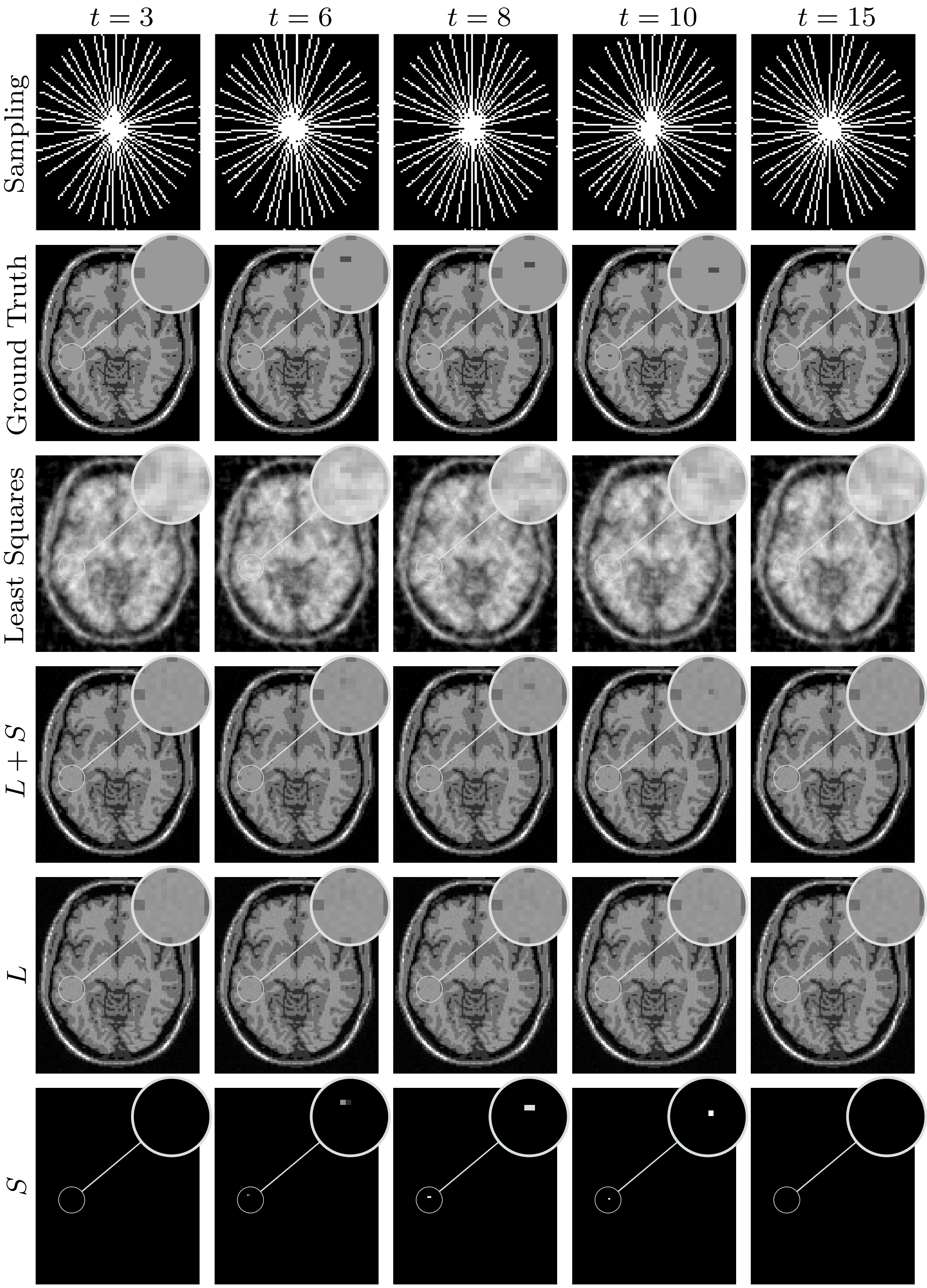


Figure 5.1: *Comparative illustration of the reconstruction results emerging from the method of least squares and the presented approach of low rank + sparse regularization with parameters* $\beta_1 = 1$ *and* $\beta_2 = 0.035$. *First row: Visualization of the (under-)sampling pattern applied to the individual time steps; Fifth and sixth row: Isolated illustration of the low rank, respectively sparse component. PSNR 'Least Squares':* 18.8738; *PSNR 'Low Rank + Sparse':* 30.1910.

duplication of this slice we then receive a constant time series of 20 frames. Manipulating this series by introducing a two pixel sized moving cell to the frames of time steps 5 to 10, this data set shall represent our ground truth. Based on this we can now extrapolate the corresponding exact k-space data by applying the Fourier transform. To mimic an undersampled data generation, we can then choose a sampling scheme which reduces the frequency data to a small fraction. In order to meet the incoherence condition between the image and frequency domain which was mentioned in Chapter 4, we here want to choose a golden ratio radial sampling simulating the scan of 20 spokes per time frame. With that we reduce the original data by approximately 82%. Further, to imitate the typical occurrence of inexact measurements we additionally introduce additive Gaussian noise with zero mean and standard deviation $\lambda = 0.035$.
With this, the constructed scenario exactly coincides with the one addressed in Section 4.3.2. Expecting a dynamic reconstruction which features a more or less constant background superposed by only a very few small moving elements suggests to solve the optimization problem in (4.7). Note that here the described setting also guarantees the requested incoherence between the low rank and the sparse component and thus makes them uniquely identifiable. With the help of the iterative method in (5.10) we are therefore able to compute the desired reconstruction. And in fact, by depicting the absolute values of the complex-valued result of this algorithmic procedure, Figure 5.1 confirms our anticipation: In comparison to a simple least squares solution, through the inclusion of a-priori knowledge on the composition of superimposing components we are able to achieve a remarkably more detailed reconstruction. While the fragmentary characteristic of the input data makes it impossible for the least squares reconstruction to feature delicate contours, the reconstruction stemming from the low rank + sparse approach succeeds well in this. Although the artificially introduced cell only measured two pixels, here the highly incomplete data was still sufficient to detect this attribute. Moreover, looking at the isolated illustrations of the low rank, respectively sparse, component, we recognize that the separation into background and motion was performed quite accurately. Consequently, with respect to neuroimmune cell imaging, applying the approach presented in Section 4.3.2 lets us profit from the beneficial side effect that the specimens of interest are automatically separated from the remainder. Additional to the improved reconstruction, this even more facilitates the evaluation of an MR scan for the radiologists and medical specialists.
In summary we can state that in settings like the presented one imposing a superposition of background and motion by solving (4.7) represents a valid tool for reconstructing highly undersampled k-space data. Partial regularization with the nuclear norm and the $1,1$-norm here allows to abbreviate the process of acquisition and/or impose a more dense temporal sampling without giving up on the quality of the reconstruction.

5.2.2 Tracking of dynamics

As already mentioned in the previous section, in the context of cell tracking, the method in (4.7) allows for more than only the efficient reconstruction of raw Fourier data. It also enables to automatically detect the contemplated objects of interest. Since this considerably simplifies

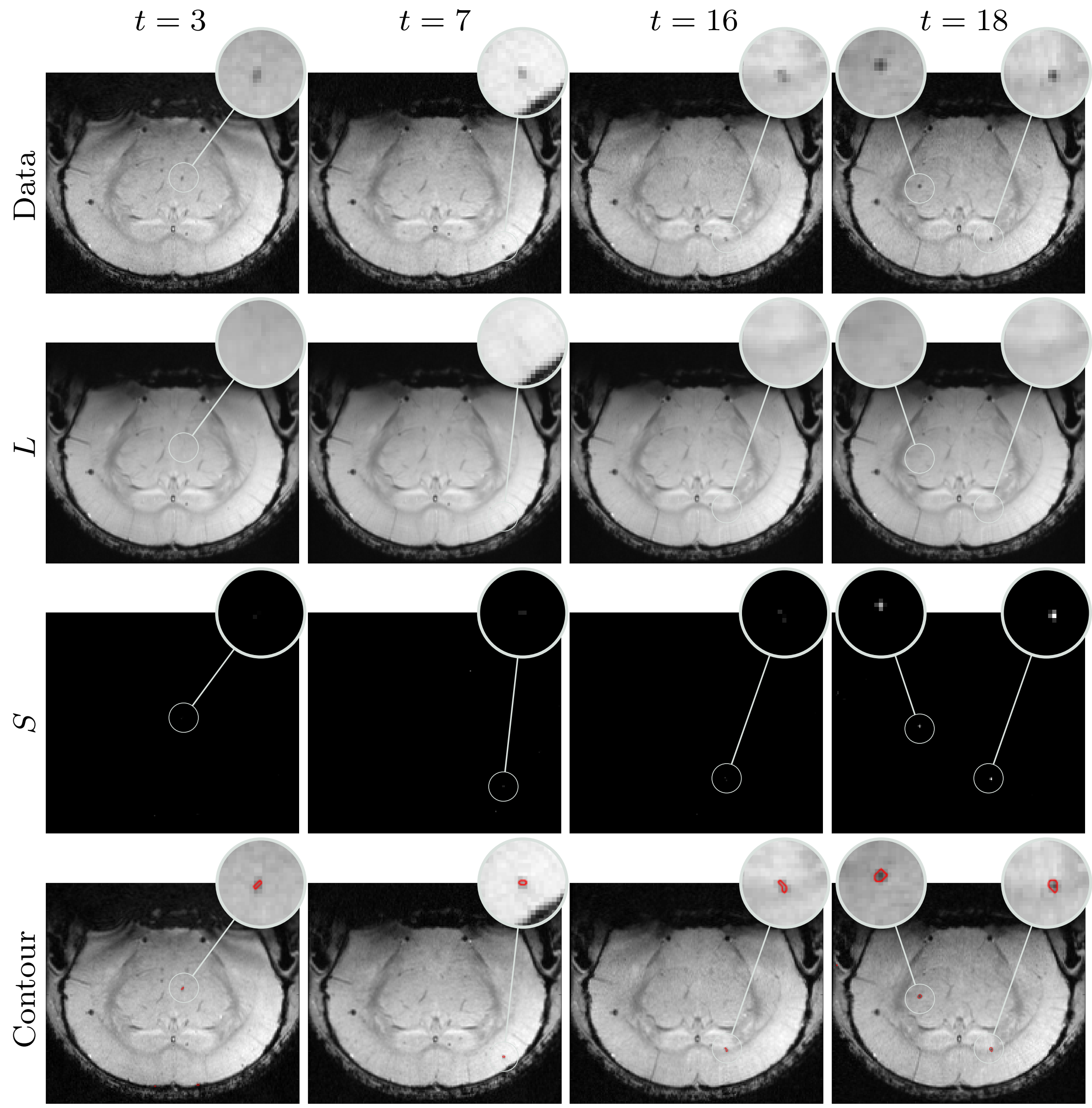


Figure 5.2: *Illustration of the cell tracking results emerging from the presented approach of low rank + sparse regularization with $K = Id$ and parameters $\beta_1 = 0.02$ and $\beta_2 = 0.00045$. Fourth row: Contour visualization of the non-zero areas of the sparse component with underlying original data.*

the work of physicians, it is advisable to check whether the low rank + sparse approach is also able to perform an a posteriori detection, i.e., if it is possible to apply the associated algorithm to data which is already reconstructed.

To do so, we want to turn toward a real dynamic MR sequence depicting neuroimmune cells in a mouse brain, which was provided to us by the Translational Research Imaging Center (TRIC) of the University Hospital Münster (see [Masthoff et al., 2019]). Here, the data which was already reconstructed via an integrated standard procedure, stems from a T2-weighted scan that was supported by the in vivo marking of the cells of interest. The full sampling of the brain slice to be imaged resulted in a time series of 20 frames, each with an acquisition time of approximately 8 min.

In order to apply the approach in (4.7) to this scenario and automatically detect the moving cells, we first have to make a small adjustment. In doing so, we especially take into account, that here the main objective of the implementation of this method is no longer the reconstruction of Fourier data. Thus, since the given data B already lies in the image space, we can choose to set the involved operator K to identity. However, before now we again consult the algorithmic scheme in (5.10) to compute the desired separation into background and motion, for this real data set we first want to make sure, that as little acquisition-induced motion as possible disturbs the tracking of the cells. Therefore, we primarily perform a very simple rigid image registration using FAIR (see [Modersitzki, 2009]). With that we can guarantee, that the images of the present data set are roughly aligned based on the underlying anatomical structure. After this preliminary step, we then can finally proceed following the iterative method presented in Section 5.1. Its outcome can be found in Figure 5.2.

Here, although the real medical data set forced us to deal with acquisition-induced shadow artifacts and a background that, despite registration, was not completely static, we can observe impressively good results. Especially in the contour plot, which highlights the areas captured by the sparse component, we see that the procedure enabled us to detect many of the small moving cells. Simultaneously, as to be expected, the low rank component depicts a far more static version of the background observed in the original data. Hence, exploiting the high linear dependence among the single frames' background and the sparse characteristic in space and time of the fast moving cells once again proved to be successful. Overall, we can thus confirm that the low rank + sparse approach also suits to retroactively track small cells in dynamic MR sequences. With that it can significantly support radiologists in the diagnostic evaluation of such data.

5.2.3 Undersampling imposing the occurrence of few smooth dynamics?

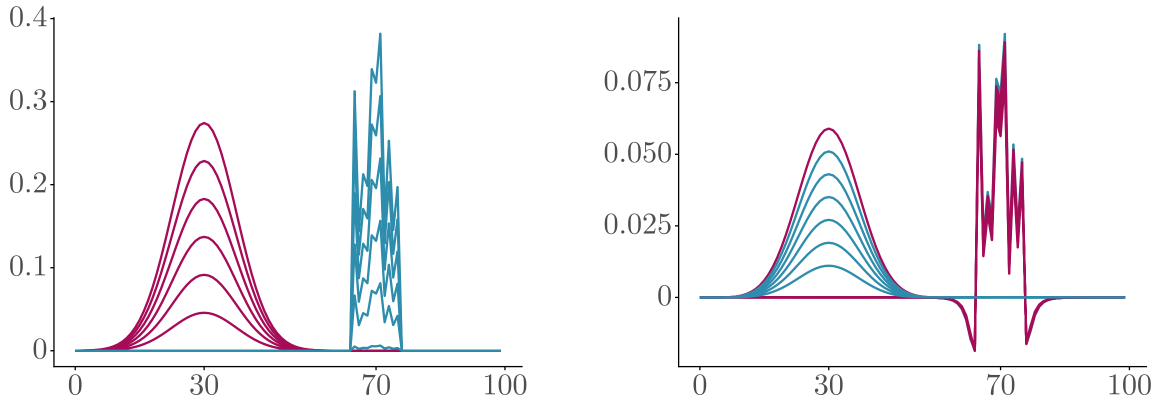
To conclude this chapter of numerical experiments on the reconstruction of dynamic MR data we lastly want to turn toward the regularization technique which involves the newly introduced generalized nuclear norm. As already pointed out in Section 4.3.3, tailored to the context of dynamic MRI we here in particular want to concentrate on the scenario in which the corresponding generalized SVD is induced by $C = I_m$ and D approximating the negative Laplacian matrix. Through this choices we hope that we can gain control over the type of dynamic which eventually ends up being featured in the computed reconstruction when limiting the rank of our solution.

In Section 5.1 we were unfortunately not able to explicitly prove that regularizing with the generalized nuclear norm boils down to a shrinkage of the generalized singular values. Instead, based on the results regarding the classical nuclear norm in Lemma 5.4, we only anticipated a similar behavior. This is why, before applying the algorithm derived in (5.15) to the reconstruction of raw MR data, we should briefly concentrate on approving our intuition in a numerical manner. Thus, before anything else, we first want to address the denoising problem in (5.11) for varying regularization parameters $\tau > 0$. Fortunately, in order to solve this problem numerically, we can recycle the iterative scheme introduced in (5.15). For this purpose the only thing left to do is to temporarily set the involved operator $\widetilde{K}$ to identity. Following this computation, we then want to calculate the respective new generalized singular values and corresponding singular vectors of our solutions in order to be able to check our conjecture. Regarding the latter, we are particularly interested in the ones representing the temporal evolution, i.e., the generalized right-singular vectors.

To get started we first want to be able to evaluate the behavior of solutions to (5.11) in a preferably simple setting. Therefore, we initially focus on a stylized minimal example: The data A we use is composed of a smooth and a non-smooth dynamic added to a 100 time frames series of zero images. During this data generation we make sure that, on a spatial basis, both dynamics can be clearly separated from each other. Moreover, also in a temporal sense we rely on the distinct differentiation of both signals. This allows us to use the analytically understood solutions to the problem in (5.6) and their classical SVDs for a comparison. That is because, in contrast to the experiments depicted in Figure 4.3 and Figure 4.4, the significantly staggered occurrence of dynamics even enables the classical SVD to feature singular vectors which separate these signals accurately from each other.

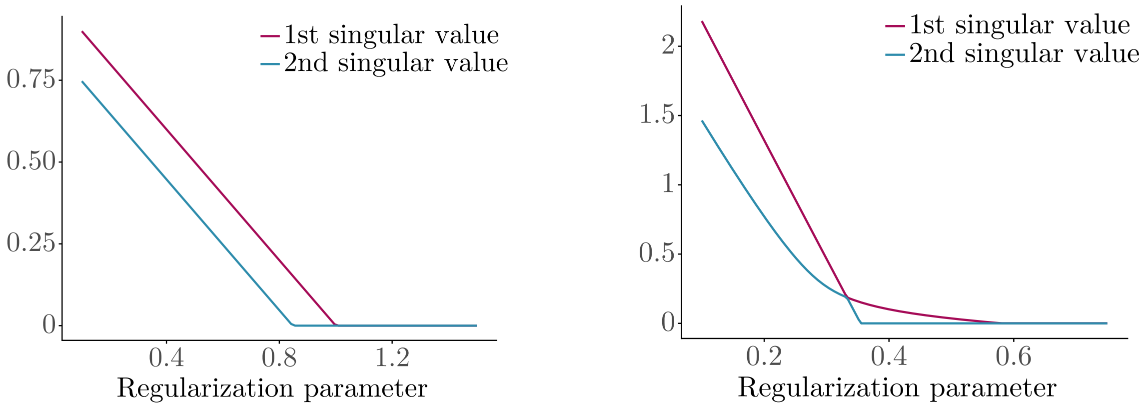
When considering the results of this comparative case along increasing regularization parameters (see Figure 5.3 (a)), we can observe the expected behavior: Through the linear shrinkage of both classical singular values the introduced signals, which get identified by the corresponding singular vectors, decrease in a uniform manner. Here, the prevailing linearity of this descent can be substantiated in Figure 5.4 (a). Depicting the isolated course of both non-zero singular values this illustration emphasizes their structure-preserving diminution and demonstrates their final vanishing behavior for big choices of τ.

Surprisingly, contemplating the same experiment with respect to the generalized nuclear norm induced by $C = I_m$ and D as in (4.10) with $\varepsilon = 10$, we can notice a completely different shrinking behavior (see Figure 5.3 (b) and Figure 5.4 (b)). While the generalized singular value corresponding to the smooth dynamic is still shrinking in a linear manner, the one corresponding to the non-smooth signal experiences a decreasingly strong shrinkage as the regularizing parameter is increasing. This suggests that the regularization with the generalized nuclear norm induces a shrinking behavior that is signal-dependent. In the present example, this has the consequence that also the dominance of the signals (here encoded by the colors red and blue) switches as the optimization in (5.11) is performed for decreasing τ. First of all, as derived in Section 4.3.3, through the use of the approximate negative Laplacian the largest singular value is assigned to the smooth dynamic. However, for parameters bigger than 0.333 this impression is shifting. Depicted through a change of color, in Figure 5.4 (b) we can observe that from this point on the singular value assigned to the non-smooth dynamic is dominant. Consulting the illustration in Figure 5.3 (b), this can be explained by the significantly stronger



(a) (b)

Figure 5.3: *Illustration of the first (red) and second (blue) singular vectors multiplied with their associated singular values after solving* (5.6) (a), *resp.* (5.11) (b), *with varying regularization parameter* $\tau \in [0,1]$, *resp.* $\tau \in [0.33, 0.35]$. *Underlying data: Two locally distinguishable dynamics introduced to a series of* 100 *images with zero background; first dynamic – Gaussian kernel over time centered in time frame 30 and with standard deviation* $\lambda = 7.5$, *second dynamic – jump function for time frames 65-75.*



(a) (b)

Figure 5.4: *Graphical illustration of the first two singular values corresponding to the experiment in Figure 5.3.*

adherence to this signal. Accordingly, we have to realize that solutions of the denoising problem in (5.11) act substantially different than the ones resulting from the corresponding problem which involves the classical nuclear norm. Contrary to our expectations, they do not reliably preserve the structure of the input data. This is especially astonishing since, respecting the rewriting options in (5.13) and (5.14), this behavior exactly coincides with the one which can be observed when applying the established concept in (4.6) to solve the inverse problem

$$TA = B$$

with smoothing operator T. Thus, although we can analytically show that in the denoising scenario the regularization with the classical nuclear norm leads to a linear shrinkage of the singular values (see Lemma 5.4), we can not guarantee a similar behavior when a smoothing forward operator is involved. Contrary to the assumptions which are commonly made, this suggests that also with respect to other operators it is not necessarily clear how regularizing with the classical nuclear norm affects the solution process. In any case, it does not seem to be advisable to extrapolate from the structure preserving shrinkage trend which can be observed and proven in the denoising context to more general scenarios which involve a true forward operator. Instead, it would be desirable to investigate more explicitly in order to understand which kind of dynamic components are primarily considered when determining low-rank solutions via minimization problems of the form (4.6). In particular, it could be interesting to grasp under which conditions regularizing with the classical nuclear norm provokes the kind of results which can for example be observed in the experiments in [Candès and Recht, 2009] and [Recht et al., 2010].

All in all, on the basis of these results, we have to admit that, even in the simple denoising setting, regularizing with the generalized nuclear norm behaves contrary to our intuition. Instead, it unfortunately obeys a scheme which is not tangible for us at the moment. For now, this makes it particularly impractical to apply the presented reconstruction technique in (4.11). However, through our studies we were able to uncover that also the well-established method in (4.6) requires further investigation. Having such investigations on the inclusion of general forward operators at hand, one could then also resume the reconstruction method introduced in Section 4.3.3.

For the moment, however, it only remains to study this approach from an analytical point of view. And indeed, this is what we want to look into in the following chapter.

6 Asymptotic behavior

Within the previous chapters we took a closer look on a selection of norms that operate on composed matrices and discussed their respective benefits in the context of discrete variational methods. As representatives of time series of images the consideration of these matrices now naturally arouses the interest in the asymptotic behavior of the studied functionals. Choosing finer and finer time steps while imposing a higher and higher image resolution makes us want to grasp how those functionals act as soon as the dimensions m and n of their domain tend to infinity.
In this chapter we thus want to figure out how the considered functionals can be translated to operate on 'more continuous' domains. For this purpose we want to employ the concept of Γ-convergence. As already pointed out in the preliminaries, provided that a sequence of functionals is equi-coercive, its application guarantees the convergence of minimizers toward minimizers. Concerning the contemplated setting it therefore promotes the emergence of limit functionals whose associated minimization problems generate solutions which are close to their high-dimensional discrete counterparts. In contrast to the application of variational methods to discrete dynamic problems with very fine temporal and spatial resolutions the determination of a Γ-limit thus allows for a very convenient analysis.

In order to face the determination of this Γ-limit in the following we first of all will figure out how a 'more continuous' domain can be understood. Moreover, we will find a general semi-discrete formulation of the previously considered energy functionals that transfers our discrete comprehension to these continuous spaces as soon as operators can be represented by a matrix. Subsequently, we will observe that the resulting sequence of functionals depending on the dimensions m and n is equi-coercive and the involved sequence of data fidelity terms converges continuously as m and n tend toward infinity. Incorporating the stability of the Γ-convergence under continuous perturbations (cf. Proposition 2.23) our analysis hence reduces to the determination of the Γ-limits of the regularizing norms. Concentrating on the special characteristics of the mixed norm, the nuclear norm and the generalized nuclear norm we conclude this chapter by examining their individual limit behavior as dimensions increase. Reassembling these single results we then found continuous counterparts of the previously introduced discrete variational problems. These maintain the existing minimizing structure and are therefore ready to be consulted for the efficient approximate solving of high-dimensional discrete problems.

6.1 Analytical study of general energy functionals

Before turning to the actual interest of this chapter – the Γ-convergence of the mixed norm, the nuclear norm and the generalized nuclear norm – we need to implement the yet contemplated energy functionals in a setting which formally allows for this limit observation. To this end we want to introduce three different continuous spaces of interest and find semi-discrete representations that are able to operate on them but still resort to the discrete concepts. Further, justifying the subsequent neglection of the data fidelity term, we want to address its continuous convergence. While doing so we additionally determine the topology with respect to which a general equi-coercivity can be achieved.

6.1.1 Formal setting

As already insinuated, motivated by our previous achievements in the discrete setting, in this chapter we strive for a translation of functionals $\mathcal{F}_\alpha : \mathbb{R}^{m\times n} \to \mathbb{R}$ of the form

$$\mathcal{F}_\alpha(A) = \frac{1}{2}||\widetilde{K}A - B_{m,n}||^2_{\omega;2,2} + \alpha\mathcal{R}(A), \tag{6.1}$$

with linear and continuous (and therefore compact) operator $\widetilde{K} : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ and regularizer $\mathcal{R} : \mathbb{R}^{m\times n} \to \mathbb{R}$ with

$$\mathcal{R}(A) \geq \beta||A||_{\omega;p,q} \quad \forall A \in \mathbb{R}^{m\times n} \tag{6.2}$$

for some $\beta > 0$, to a setting which in some sense is 'more continuous'. This gain of continuity can be understood in three different ways: As a first attempt one could address the transition to a space-continuous setting in which discrete time intervals are preserved. From an applied point of view this becomes relevant for example in the context of cardiac MRI. Intending to image an ever-adapting organ the ECG-supported gating technique, which allows for the collection of measurements during short periods of cardiac quiescence, proved to be a useful tool. Hence, while an increasing space-resolution contributes to the precision of this imaging method, sticking to a special discrete set of time steps is crucial and should be integrated into the limit observation. From an interpretive point of view the contrary scenario in which one is interested in studying the behavior of the respective functionals in a time-continuous though space-discrete setting is less intuitive. Nevertheless, for the sake of completeness we also want to discuss this second approach of continuity. Obviously, the third option of introducing more continuity to problems like the one in (6.1) is the observation of their behavior as soon as both, the temporal as well as the spatial dimension variable, tend to infinity. This then represents the target scenario of nearly all prevalent dynamic imaging techniques. Taken all together this means that within the scope of this chapter we are interested in the following three scenarios:

(I) *Continuity in space*, i.e. consideration of the dimension pair $(m_\mu, n_\mu) \coloneqq (m_\mu, n) \in \mathbb{N} \times \mathbb{N}$ with $m_\mu \to \infty$ as $\mu \to \infty$,

(II) *Continuity in time*, i.e. consideration of the dimension pair $(m_\mu, n_\mu) \coloneqq (m, n_\mu) \in \mathbb{N} \times \mathbb{N}$ with $n_\mu \to \infty$ as $\mu \to \infty$,

(III) *Continuity in space and time*, i.e. consideration of the dimension pair $(m_\mu, n_\mu) \in \mathbb{N} \times \mathbb{N}$ with $m_\mu, n_\mu \to \infty$ as $\mu \to \infty$.

In order to put these approaches in more formal words we want to agree on dealing with the open subsets $\Sigma \subset \mathbb{R}^d$ and $\Omega \subset \mathbb{R}^{d'}$. Additionally, respecting the analogy between matrices representing linear operators which map vectors on vectors and general operators which handle functions, we can then perceive that the previously mentioned scenarios aim for energy functionals that are able to operate on a space $\mathcal{T}$, which can be identified with:

(I) $$\mathcal{I}^{q,p}(\mathbb{R}^n, \Sigma) \coloneqq \left\{ T : \mathbb{R}^n \to L^p(\Sigma) \middle| (Ty)(s) = \sum_{j=1}^{n} y_j t_j(s) \text{ for } t \in L^p(\Sigma; \mathbb{R}^n) \right\},$$

(II) $$\mathcal{I}^{q,p}(\Omega, \mathbb{R}^m) \coloneqq \left\{ T : L^{q'}(\Omega) \to \mathbb{R}^m \middle| (Ty)_i = \int_\Omega t_i(r) y(r) \,\mathrm{d}r \text{ for } t \in L^q(\Omega; \mathbb{R}^m) \right\},$$

(III) $$\mathcal{I}^{q,p}(\Omega, \Sigma) \coloneqq \left\{ T : L^{q'}(\Omega) \to L^p(\Sigma) \middle| (Ty)(s) = \int_\Omega t(s,r) y(r) \,\mathrm{d}r \text{ for } t \in L^{p,q}(\Sigma \times \Omega) \right\}.$$

Note that within all three different definitions the operators are uniquely defined by an associated element t. In order to refer to this element in the following more easily, regardless of the considered continuity scenario, we generally want to assume that $t \in \mathfrak{T}$. Thus, we can understand $\mathfrak{T}$ as a placeholder for the spaces $L^p(\Sigma, \mathbb{R}^n)$, $L^q(\Omega; \mathbb{R}^m)$ and $L^{p,q}(\Sigma \times \Omega)$. Speaking of notations that will facilitate our further analysis, we also need to mention that, when making statements which equally apply to all spaces, i.e. to $\mathcal{I}^{q,p}(\mathbb{R}^n, \Sigma)$, $\mathcal{I}^{q,p}(\Omega, \mathbb{R}^m)$ *and* $\mathcal{I}^{q,p}(\Sigma, \Omega)$, we want to use the general notation $\mathcal{T}$ to refer to them. Since the case in which $p = q = 2$ will play a special role we moreover want to specify this notation by using $\mathcal{T}_2$ as soon as p and q are fixed in that way. Beyond that, concerning the limit behavior we generally want to restrict ourselves to considering $\mu \to \infty$. Thereby we leave it up to the reader to transfer the behavior of μ to the behavior of the respective pair (m_μ, n_μ) and the corresponding continuity scenario in (I), (II) or (III).

With these definitions and notations at hand we first of all find that in all three cases the space of matrices $\mathbb{R}^{m \times n}$ can naturally be embedded into $\mathcal{T}$. Assuming that Σ and Ω are equipped with partitions $\mathcal{S}_m = \{\Sigma_1^m, \dots \Sigma_m^m\}$ and $\mathcal{O}_n = \{\Omega_1^n, \dots, \Omega_n^n\}$ which divide them into m, respectively n, pairwise disjoint subsets that obey

$$|\Sigma| = \sum_{i=1}^{m} |\Sigma_i^m|, \qquad |\Omega| = \sum_{j=1}^{n} |\Omega_j^n|,$$

we can specify this embedding via the introduction of the operator $E : \mathbb{R}^{m \times n} \to \mathcal{T}$.

Definition 6.1. Let $\mathcal{T}$ be identified with the continuous space corresponding to scenario (I), (II) or (III). Then, depending on the considered scenario, the embedding operator $E : \mathbb{R}^{m\times n} \to \mathcal{T}$ is defined as

$$\text{(I)}\quad [(EA)(y)](s) := \sum_{j=1}^{n} y_j \left(\sum_{i=1}^{m} A_{ij}\chi_{\Sigma_i^m}(s) \right) \quad \forall y \in \mathbb{R}^n, s \in \Sigma,$$

$$\text{(II)}\quad [(EA)(y)]_i := \int_\Omega \left(\sum_{j=1}^{n} A_{ij}\chi_{\Omega_j^n}(r) \right) y(r)\,\mathrm{d}r \quad \forall y \in L^{q'}(\Omega), i \in \{1,\dots m\},$$

$$\text{(III)}\quad [(EA)(y)](s) := \int_\Omega \left(\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}\chi_{\Sigma_i^m}(s)\chi_{\Omega_j^n}(r) \right) y(r)\,\mathrm{d}r \quad \forall y \in L^{q'}(\Omega), s \in \Sigma.$$

With the help of this embedding we now want to consider the sought-after continuous translation of functionals of the form (6.1) reasonable as soon as their evaluation for discretely representable operators, i.e. for operators $T = EA$ with $A \in \mathbb{R}^{m\times n}$, coincides with $\mathcal{F}_\alpha(A)$. Following this claim we can characterize a functional which naturally lifts our understanding of $\mathcal{F}_\alpha$ to a semi-discrete level. To do so we define $\mathcal{E}_\alpha^{m,n} : \mathcal{T} \to \overline{\mathbb{R}}$ via

$$\mathcal{E}_\alpha^{m,n}(T) := \begin{cases} \mathcal{F}_\alpha(A), & T = EA \text{ for some } A \in \mathbb{R}^{m\times n}, \\ \infty, & \text{else.} \end{cases} \tag{6.3}$$

Trying to minimize this functional over $\mathcal{T}$ we thus obtain a series of problems whose elements are equivalent to the minimization of $\mathcal{F}_\alpha$ over the respective space of matrices.

Now the objective of the following is to determine the limit of the resulting sequence of functionals $(\mathcal{E}_\alpha^{m_\mu,n_\mu})_{\mu\in\mathbb{N}}$ with respect to the three desired continuity scenarios. In order to maintain the structure of the single minimization problems we want to pursue this goal using the concepts of Γ-convergence. Therefore, we start by carefully examining how the operation of $\mathcal{E}_\alpha^{m_\mu,n_\mu}$ on its domain $\mathcal{T}$ can be understood. To do so we first of all want to concentrate on the space of all discretely representable operators, i.e. on

$$\mathcal{T}^{m_\mu,n_\mu} := \left\{ T \in \mathcal{T} \,|\, T = EA \text{ for some } A \in \mathbb{R}^{m_\mu\times n_\mu} \right\},$$

and note the following property.

Remark 6.2. Considering an operator $T \in \mathcal{T}^{m_\mu,n_\mu} \subset \mathcal{T}$ with $T = EA$ for $A \in \mathbb{R}^{m_\mu\times n_\mu}$ we can find a norm which resorts exclusively to the structures in $\mathcal{T}$ but simultaneously coincides with the weighted mixed norm of A. In order to characterize this norm more explicitly we want to cater to the single specifications of $\mathcal{T}$:

- $\mathcal{T} = \mathcal{I}^{q,p}(\mathbb{R}^n, \Sigma)$: In this scenario T is element of $\mathcal{T}$ with

$$t = \left(\sum_{i=1}^{m} A_{ij} \chi_{\Sigma_i^m} \right)_{j=1,\dots,n} \in L^p(\Sigma; \mathbb{R}^n).$$

Hence, due to the pairwise disjointness of the Σ_i^m and the range of the $\chi_{\Sigma_i^m}$ we can deduce that for $\omega_{ij} := |\Sigma_i^m|^{\frac{q}{p}}$

$$\begin{aligned} ||A||_{\omega;p,q} &= \left(\sum_{i=1}^{m} \left(\sum_{j=1}^{n} |\Sigma_i^m|^{\frac{q}{p}} |A_{ij}|^q \right)^{\frac{p}{q}} \right)^{\frac{1}{p}} \\ &= \left(\int_\Sigma \sum_{i=1}^{m} \left(\sum_{j=1}^{n} |A_{ij}|^q \right)^{\frac{p}{q}} \chi_{\Sigma_i^m}(s) \, \mathrm{d}s \right)^{\frac{1}{p}} \\ &= \left(\int_\Sigma \left(\sum_{j=1}^{n} \sum_{i=1}^{m} |A_{ij}|^q \chi_{\Sigma_i^m}(s) \right)^{\frac{p}{q}} \mathrm{d}s \right)^{\frac{1}{p}} \\ &= \left(\int_\Sigma \left(\sum_{j=1}^{n} \left| \sum_{i=1}^{m} A_{ij} \chi_{\Sigma_i^m}(s) \right|^q \right)^{\frac{p}{q}} \mathrm{d}s \right)^{\frac{1}{p}} \\ &= \left(\int_\Sigma \left(\sum_{j=1}^{n} |t_j(s)|^q \right)^{\frac{p}{q}} \mathrm{d}s \right)^{\frac{1}{p}} \end{aligned} \tag{6.4}$$

applies.

- $\mathcal{T} = \mathcal{I}^{q,p}(\Omega, \mathbb{R}^m)$: In this scenario T is element of $\mathcal{T}$ with

$$t = \left(\sum_{j=1}^{n} A_{ij} \chi_{\Omega_j^n} \right)_{i=1,\dots,m} \in L^q(\Omega; \mathbb{R}^m).$$

Hence, incorporating the disjointness of the Ω_j^n we deduce similar to the previous case that for $\omega_{ij} := |\Omega_j^n|$

$$||A||_{\omega;p,q} = \left(\sum_{i=1}^{m} \left(\int_\Omega \sum_{j=1}^{n} |A_{ij}|^q \chi_{\Omega_j^n}(r) \, \mathrm{d}r \right)^{\frac{p}{q}} \right)^{\frac{1}{p}}$$

$$= \left(\sum_{i=1}^{m} \left(\int_\Omega \left| \sum_{j=1}^{n} A_{ij} \chi_{\Omega_j^n}(r) \right|^q \mathrm{d}r \right)^{\frac{p}{q}} \right)^{\frac{1}{p}}$$

$$= \left(\sum_{i=1}^{m} ||t_i||_{L^q(\Omega)}^p \right)^{\frac{1}{p}} \tag{6.5}$$

holds true.

- $\mathcal{T} = \mathcal{I}^{q,p}(\Omega, \Sigma)$: In this scenario T is element of $\mathcal{T}$ with

$$t = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} \chi_{\Sigma_i^m} \chi_{\Omega_j^n} \in L^{p,q}(\Sigma \times \Omega).$$

Hence, as a combination of the previous two cases we analogously validate that

$$||A||_{\omega;p,q} = ||t||_{L^{p,q}(\Sigma\times\Omega)} \tag{6.6}$$

applies for $\omega_{ij} := |\Sigma_i^m|^{\frac{q}{p}} |\Omega_j^n|$.

Consequently, using the on $t \in L^p(\Sigma; \mathbb{R}^n)$, $t \in L^q(\Omega; \mathbb{R}^m)$ or $t \in L^{p,q}(\Sigma \times \Omega)$ dependent expressions in (6.4), (6.5) and (6.6) it is possible to naturally define an operator norm which turns $\mathcal{T}^{m_\mu, n_\mu}$ into a normed space.

In the course of this chapter we will see that the candidates derived in Remark 6.2 are also good choices for a respective norm on the entire extent of $\mathcal{T}$. This is why we also want to introduce them in a formal way.

Definition 6.3. Let $\mathcal{I}^{q,p}(\mathbb{R}^n, \Sigma)$, $\mathcal{I}^{q,p}(\Omega, \mathbb{R}^m)$ and $\mathcal{I}^{q,p}(\Omega, \Sigma)$ be defined according to the scenarios in (I), (II) or (III). Then, through

$$\text{(I)} \quad ||T||_{\mathcal{I}^{q,p}(\mathbb{R}^n,\Sigma)} := \left(\int_\Sigma \left(\sum_{j=1}^{n} |t_j(s)|^q \right)^{\frac{p}{q}} \mathrm{d}s \right)^{\frac{1}{p}},$$

$$\text{(II)} \quad ||T||_{\mathcal{I}^{q,p}(\Omega,\mathbb{R}^m)} := \left(\sum_{i=1}^{m} ||t_i||_{L^q(\Omega)}^p \right)^{\frac{1}{p}},$$

$$\text{(III)} \quad ||T||_{\mathcal{I}^{q,p}(\Omega,\Sigma)} := ||t||_{L^{p,q}(\Sigma\times\Omega)}$$

we define the normed spaces $\big(\mathcal{I}^{q,p}(\mathbb{R}^n,\Sigma), ||\,.\,||_{\mathcal{I}^{q,p}(\mathbb{R}^n,\Sigma)}\big)$, $\big(\mathcal{I}^{q,p}(\Omega,\mathbb{R}^m), ||\,.\,||_{\mathcal{I}^{q,p}(\Omega,\mathbb{R}^m)}\big)$ and $\big(\mathcal{I}^{q,p}(\Omega,\Sigma), ||\,.\,||_{\mathcal{I}^{q,p}(\Omega,\Sigma)}\big)$.

Note that, since these norms resort to well-known norms on $\mathbb{R}^n$, $\mathbb{R}^m$, $L^p(\Sigma)$, $L^q(\Omega)$ and $L^{p,q}(\Sigma\times\Omega)$, their corresponding normed spaces are even complete, i.e. they are Banach spaces. In the special case in which $p = q = 2$ we can go even further.

Remark 6.4. Let $p = q = 2$. Then, we can extend the characterizations in Definition 6.3 via the introduction of the following inner product:

$$\langle T, L\rangle_{\mathcal{T}_2} := \langle t, \ell\rangle_{\mathfrak{T}_2}.$$

Through this, $(\mathcal{T}_2, \langle\,.\,,\,.\,\rangle_{\mathcal{T}_2})$ defines for all three characteristics of $\mathcal{T}_2$ a Hilbert space.

With these insights we now want to grasp how the compact operator used in formulation (6.1) can be extrapolated to operate on a discretely representable operator $T \in \mathcal{T}^{m_\mu,n_\mu}$. To do so we first of all need to notice that within the discrete formulation the exact definition of this operator adapts to the respective dimensions $m, n \in \mathbb{N}$. In order to incorporate that in the present setting we are considering a sequence of functionals for which, depending on the desired continuity scenario, one or both of these parameters vary, from now on we want to work with an extended version. Combining all the single definitions, this extended version should then be able to deal with different sizes of matrices. Therefore, without relabeling, we now assume $\widetilde{K}$ to work as an operator between $\bigcup_{\mu\in\mathbb{N}} \mathbb{R}^{m_\mu\times n_\mu}$ and $\bigcup_{\mu\in\mathbb{N}} \mathbb{R}^{m_\mu\times n_\mu}$. Based on this we now want to find an equivalent definition which operates directly on $\bigcup_{\mu\in\mathbb{N}} \mathcal{T}^{m_\mu,n_\mu}$. With this objective in mind we characterize $\mathring{K} : \bigcup_{\mu\in\mathbb{N}} \mathcal{T}^{m_\mu,n_\mu} \to \bigcup_{\mu\in\mathbb{N}} \mathcal{T}_2^{m_\mu,n_\mu}$ via

$$\mathring{K}T := E(KA).$$

Interestingly, due to the coincidence of $\mathcal{T}$ with the closure of $\bigcup_{\mu\in\mathbb{N}} \mathcal{T}^{m_\mu,n_\mu}$ regarding the respective norm (cf. Definition 6.3) this operator can be continuously extended to $\mathcal{T}$. In the following we want to presume that $\widetilde{K}$ was already chosen in a way such that this expansion $\overline{K} : \mathcal{T} \to \mathcal{T}_2$ also adopts its compactness.

With the definition of this compact operator we are now able to specify the representation of $\mathcal{E}_\alpha^{m_\mu,n_\mu} : \mathcal{T} \to \overline{\mathbb{R}}$ with respect to the data fidelity term. Construing $B_{m_\mu,n_\mu} \in \mathbb{R}^{m_\mu\times n_\mu}$ in formulation (6.1) as the operator $EB_{m_\mu,n_\mu} \in \mathcal{T}_2$ we realize that

$$\mathcal{E}_\alpha^{m_\mu,n_\mu} = \mathcal{D}_{B_{m_\mu,n_\mu}} + \alpha\mathcal{R}^{m_\mu,n_\mu} \tag{6.7}$$

with $\mathcal{D}_{B_{m_\mu,n_\mu}} : \mathcal{T} \to \mathbb{R}$, $\mathcal{R}^{m_\mu,n_\mu} : \mathcal{T} \to \overline{\mathbb{R}}$ defined by

$$\mathcal{D}_{B_{m_\mu,n_\mu}}(T) := \frac{1}{2}||\overline{K}T - EB_{m_\mu,n_\mu}||^2_{\mathcal{T}_2} \tag{6.8}$$

$$\mathcal{R}^{m_\mu,n_\mu}(T) := \begin{cases} \mathcal{R}(A), & \text{if } T = EA \text{ for some } A \in \mathbb{R}^{m_\mu\times n_\mu}, \\ \infty, & \text{else.} \end{cases} \tag{6.9}$$

Note that in this representation the dependence of the data fidelity term on the dimensions m_μ and n_μ relates exclusively to the operator which is induced by $B_{m_\mu,n_\mu} \in \mathbb{R}^{m_\mu \times n_\mu}$. In contrast to the definition in (6.3), neither the norm on $\mathcal{T}_2$ nor the compact operator $\overline{K}$ distinguishes between operators which can be characterized by a matrix and those that do not own a characterization like that. As a consequence, $\mathcal{D}_{B_{m_\mu,n_\mu}}$ is not only able to operate properly on $\mathcal{T}^{m_\mu,n_\mu}$, but also assigns meaningful and finite values to general operators in $\mathcal{T}$.
Within the ensuing subsections we now want to show that, given a sequence $(B_{m_\mu,n_\mu})_{\mu\in\mathbb{N}} \subset \bigcup_{\mu\in\mathbb{N}} \mathbb{R}^{m_\mu\times n_\mu}$ whose corresponding sequence of operators $(EB_{m_\mu,n_\mu})_{\mu\in\mathbb{N}} \subset \mathcal{T}_2$ converges in norm toward an operator $B \in \mathcal{T}_2$, i.e. for which

$$\lim_{\mu\to\infty} ||EB_{m_\mu,n_\mu} - B||_{\mathcal{T}_2} = 0$$

applies, these values are even meaningful in the context of Γ-convergence and it suffices to exclusively examine the Γ-limits of the regularizing parts in $\mathcal{E}_\alpha^{m_\mu,n_\mu}$. For this conclusion we make use of Proposition 2.23: By proving that $(\mathcal{D}_{B_{m_\mu,n_\mu}})_{\mu\in\mathbb{N}}$ is a continuously convergent sequence of functionals we can guarantee that for its limit $\mathcal{D}_B : \mathcal{T} \to \mathbb{R}$

$$\Gamma\text{-}\lim_{\mu\to\infty} \mathcal{E}_\alpha^{m_\mu,n_\mu} = \mathcal{D}_B + \alpha \; \Gamma\text{-}\lim_{\mu\to\infty} \mathcal{R}^{m_\mu,n_\mu}$$

applies.

6.1.2 Coercivity

Aiming for a Γ-limit that maintains the minimizing structure, i.e. whose minimizer coincides with the limit of the sequence of minimizers of $\mathcal{E}_\alpha^{m_\mu,n_\mu}$, the topology which allows for the equi-coercivity of $(\mathcal{E}_\alpha^{m_\mu,n_\mu})_{\mu\in\mathbb{N}}$ sets the target (cf. Theorem 2.20). Thus, in order to appraise with respect to which type of topology on $\mathcal{T}$ the continuous convergence of $(\mathcal{D}_{B_{m_\mu,n_\mu}})_{\mu\in\mathbb{N}}$ should be striven for, we first want to discuss this coercivity.

Proposition 6.5. Let (m_μ, n_μ) and $\mathcal{T}$ be defined according to the scenario in (I), (II) or (III). Suppose that $\mathcal{E}_\alpha^{m_\mu,n_\mu} : \mathcal{T} \to \overline{\mathbb{R}}$ is defined as in (6.7) and that the involved regularizer $\mathcal{R}$ fulfills the constraint in (6.2). Then the sequence of functionals $(\mathcal{E}_\alpha^{m_\mu,n_\mu})_{\mu\in\mathbb{N}}$ is equi-coercive with respect to the weak operator topology on $\mathcal{T}$.

Before proving this proposition we first need to understand how the weak convergence of a sequence of integral operators in $\mathcal{T}$ relates to the associated elements in $\mathfrak{T}$. Therefore, we formulate the following lemma.

Lemma 6.6. Suppose that $\mathcal{T}$ and $\mathfrak{T}$ are defined according to the scenario in (I), (II) or (III). Let furthermore $(T_\gamma)_{\gamma\in\mathbb{N}} \subset \mathcal{T}$ be a sequence of integral operators and $(t_\gamma)_{\gamma\in\mathbb{N}} \subset \mathfrak{T}$ be the sequence of associated elements in $\mathfrak{T}$ that determine the operator. Then the following two assertions are equivalent:

(i) $(T_\gamma)_{\gamma\in\mathbb{N}}$ converges with respect to the weak operator topology on $\mathcal{T}$,

(ii) $(t_\gamma)_{\gamma\in\mathbb{N}}$ converges with respect to the weak topology on $\mathfrak{T}$.

Since the proof of this lemma varies only very little with respect to the considered space of integral operators we limit ourselves to exemplifying the equivalence of the weak convergence in $\mathcal{I}^{q,p}(\mathbb{R}^n,\Sigma)$ and $L^p(\Sigma;\mathbb{R}^n)$. The other two statements can be inferred accordingly.

Proof.

(ii) ⇒ (i)
Let $(t_\gamma)_{\gamma\in\mathbb{N}}$ be a weakly convergent sequence in $L^p(\Sigma;\mathbb{R}^n)$. Then due to Corollary 2.3 there exists $t_* \in L^p(\Sigma;\mathbb{R}^n)$ such that

$$\lim_{\gamma\to\infty}\sum_{j=1}^{n}\int_\Sigma (t_\gamma)_j(s)t'_j(s)\,\mathrm{d}s = \sum_{j=1}^{n}\int_\Sigma (t_*)_j(s)t'_j(s)\,\mathrm{d}s$$

holds for all $t' \in L^{p'}(\Sigma)^n$. Hence, we can deduce

$$\lim_{\gamma\to\infty}\int_\Sigma (T_\gamma y)(s)x(s)\,\mathrm{d}s = \lim_{\gamma\to\infty}\sum_{j=1}^{n}\int_\Sigma (t_\gamma)_j(s)\,(y_j\,x(s))\,\mathrm{d}s = \sum_{j=1}^{n}\int_\Sigma (t_*)_j(s)\,(y_j\,x(s))\,\mathrm{d}s$$

for all $y \in \mathbb{R}^n$, $x \in L^{p'}(\Sigma)$ which implies the weak convergence of $(T_\gamma)_{\gamma\in\mathbb{N}}$ toward $T_* \in \mathcal{I}^{q,p}(\mathbb{R}^n,\Sigma)$ with

$$(T_*y)(s) := \sum_{j=1}^{n} y_j(t_*)_j(s).$$

(i) ⇒ (ii)
Suppose that $(T_\gamma)_{\gamma\in\mathbb{N}}$ converges weakly toward T_* with respect to the operator topology on $\mathcal{I}^{q,p}(\mathbb{R}^n,\Sigma)$. Then

$$\lim_{\gamma\to\infty}\int_\Sigma\sum_{j=1}^{n} y_j(t_\gamma)_j(s)\,x(s)\,\mathrm{d}s = \int_\Sigma (T_*y)(s)x(s)\,\mathrm{d}s$$

applies for all $y \in \mathbb{R}^n$, $x \in L^{p'}(\Sigma)$ and the associated sequence $(t_\gamma)_{\gamma\in\mathbb{N}} \subset L^p(\Sigma;\mathbb{R}^n)$ is bounded. Now considering the reflexivity of $L^p(\Sigma;\mathbb{R}^n)$ and applying Banach-Alaoglu it follows that $(t_\gamma)_{\gamma\in\mathbb{N}}$ has a weakly convergent subsequence $(t_{\gamma_\xi})_{\xi\in\mathbb{N}}$ whose limit $\bar{t}$ through Corollary 2.3 lies again in $L^p(\Sigma;\mathbb{R}^n)$. Following the argumentation in the first part of the proof this indicates the weak convergence of the implicitly defined subsequence of corresponding integral operators $(T_{\gamma_\xi})_{\xi\in\mathbb{N}}$ toward $\overline{T} \in \mathcal{I}^{q,p}(\mathbb{R}^n,\Sigma)$ with

$$(\overline{T}y)(s) := \sum_{j=1}^{n} y_j\bar{t}_j(s).$$

Beyond that, respecting the uniqueness of the weak convergence we obtain that $\overline{T}$ and T_* coincide and thus $T_* \in \mathcal{I}^{q,p}(\mathbb{R}^n, \Sigma)$.
Now let $t_* \in L^p(\Sigma; \mathbb{R}^n)$ denote the element which defines T_*. Then due to $T_\gamma \rightharpoonup T_*$

$$\lim_{\gamma\to\infty} \sum_{j=1}^{n} \int_\Sigma y_j (t_\gamma)_j(s)\, x(s)\, \mathrm{d}s = \sum_{j=1}^{n} \int_\Sigma y_j (t_*)_j(s)\, x(s)\, \mathrm{d}s \tag{6.10}$$

holds true for all $x \in L^{p'}(\Sigma)$, $y \in \mathbb{R}^n$. In particular this statement is valid for $y = e_j$ with $j \in \{1, \ldots, n\}$ and all elements x of any basis of $L^{p'}(\Sigma)$ whose combination through $(y_j\, x(s))_{j=1,\ldots,n}$ forms a basis in $L^{p'}(\Sigma; \mathbb{R}^n)$. Hence, we can infer from (6.10) the applicability of

$$\lim_{\gamma\to\infty} \sum_{j=1}^{n} \int_\Sigma (t_\gamma)_j(s)(t')_j(s)\, \mathrm{d}s = \sum_{j=1}^{n} \int_\Sigma (t_*)_j(s)(t')_j(s)\, \mathrm{d}s$$

for every element t' of this basis, which already implies the validity of the statement for any $t' \in L^{p'}(\Sigma; \mathbb{R}^n)$.

□

Note that according to the proof of this lemma the limits T_* and t_* of the bijectively related sequences $(T_\gamma)_{\gamma\in\mathbb{N}} \subset \mathcal{I}^{q,p}(\mathbb{R}^n, \Sigma)$ and $(t_\gamma)_{\gamma\in\mathbb{N}} \subset L^p(\Sigma; \mathbb{R}^n)$ also correspond to each other in the expected way. This limit behavior can of course be transferred to sequences in $\mathcal{I}^{q,p}(\Omega, \mathbb{R}^m)$ or $\mathcal{I}^{q,p}(\Omega, \Sigma)$ and their corresponding sequences of defining elements in $L^q(\Omega; \mathbb{R}^m)$ or $L^{p,q}(\Sigma \times \Omega)$. With the equivalence of weak convergences at hand we are now able to proof Proposition 6.5.

Proof of Proposition 6.5. Let $(T_\mu)_{\mu\in\mathbb{N}} \subset \mathcal{T}$ be an arbitrary sequence that fulfills

$$\sup_{\mu\in\mathbb{N}} \mathcal{E}_\alpha^{m_\mu,n_\mu}(T_\mu) < \infty.$$

Then all elements T_μ of this sequence can be represented by a matrix $A_{m_\mu,n_\mu} \in \mathbb{R}^{m_\mu \times n_\mu}$, and there exists a constant $C > 0$, such that

$$C \geq \mathcal{E}_\alpha^{m_\mu,n_\mu}(T_\mu) \quad \forall \mu \in \mathbb{N}.$$

Respecting that $\mathcal{D}_{B_{m_\mu,n_\mu}}$ is strictly positive and that the discrete regularizer obeys $\mathcal{R} \geq \beta ||.||_{\omega;p,q}$ we can deduce together with the identities in Remark 6.2 that for all $\mu \in \mathbb{N}$

$$\begin{aligned} C &\geq \mathcal{D}_{B_{m_\mu,n_\mu}}(T_\mu) + \mathcal{R}(A_{m_\mu,n_\mu}) \\ &\geq \beta ||T_\mu||_{\mathcal{T}}. \end{aligned}$$

Hence, $(T_\mu)_{\mu\in\mathbb{N}}$ is bounded in norm by $\frac{C}{\beta}$. Dependent on the different characteristics of $\mathcal{T}$ and its associated norm we can infer:

- for $\mathcal{T} = \mathcal{I}^{q,p}(\mathbb{R}^n, \Sigma)$ and $(T_\mu)_{\mu\in\mathbb{N}} \subset \mathcal{I}^{q,p}(\mathbb{R}^n, \Sigma)$ with corresponding sequence $(t_\mu)_{\mu\in\mathbb{N}} \subset L^p(\Sigma; \mathbb{R}^n)$ it holds due to the norm equivalence in finite dimensions that there exist constants $\overline{C} > 0$, $\widetilde{C} > 0$ such that for all $\mu \in \mathbb{N}$

$$\begin{aligned}
\frac{C}{\beta} &\geq ||T_\mu||_{\mathcal{I}^{q,p}(\mathbb{R}^n,\Sigma)} \\
&= \left(\int_\Sigma \left(\sum_{j=1}^n |(t_\mu)_j(s)|^q \right)^{\frac{p}{q}} \mathrm{d}s \right)^{\frac{1}{p}} \\
&\geq \left(\int_\Sigma \left(\overline{C} \left(\sum_{j=1}^n |(t_\mu)_j(s)|^p \right)^{\frac{1}{p}} \right)^{p} \mathrm{d}s \right)^{\frac{1}{p}} \\
&= \overline{C} \left(\sum_{j=1}^n ||(t_\mu)_j||^p_{L^p(\Sigma)} \right)^{\frac{1}{p}} \\
&\geq \overline{C}\widetilde{C} ||t_\mu||_{L^p(\Sigma;\mathbb{R}^n)}
\end{aligned}$$

- for $\mathcal{T} = \mathcal{I}^{q,p}(\Omega, \mathbb{R}^m)$ and $(T_\mu)_{\mu\in\mathbb{N}} \subset \mathcal{I}^{q,p}(\Omega, \mathbb{R}^m)$ with corresponding sequence $(t_\mu)_{\mu\in\mathbb{N}} \subset L^q(\Omega; \mathbb{R}^m)$ it holds due to the norm equivalence in finite dimensions that there exists a constant $\widehat{C} > 0$ such that for all $\mu \in \mathbb{N}$

$$\frac{C}{\beta} \geq ||T_\mu||_{\mathcal{I}^{q,p}(\Omega,\mathbb{R}^m)} = \left(\sum_{i=1}^m ||(t_\mu)_i||^p_{L^q(\Omega)} \right)^{\frac{1}{p}} \geq \widehat{C} ||t_\mu||_{L^q(\Omega;\mathbb{R}^m)}.$$

- for $\mathcal{T} = \mathcal{I}^{q,p}(\Omega, \Sigma)$ and $(T_\mu)_{\mu\in\mathbb{N}} \subset \mathcal{I}^{q,p}(\Omega, \Sigma)$ with corresponding sequence of integral kernels $(t_\mu)_{\mu\in\mathbb{N}} \subset L^{p,q}(\Sigma \times \Omega)$ it immediately holds that for all $\mu \in \mathbb{N}$

$$\frac{C}{\beta} \geq ||T_\mu||_{\mathcal{I}^{q,p}(\Omega,\Sigma)} = ||t_\mu||_{L^{p,q}(\Sigma\times\Omega)}.$$

Thus, in all three cases the boundedness of $(T_\mu)_{\mu\in\mathbb{N}}$ implies the boundedness of the sequence of respective associated elements in $L^p(\Sigma; \mathbb{R}^n)$, $L^q(\Omega; \mathbb{R}^m)$ or $L^{p,q}(\Sigma \times \Omega)$. Taking into account the reflexivity of these spaces (cf. Remark 2.2, Theorem 2.6) and following Banach-Alaoglu this in turn induces their convergence, up to subsequences, with respect to the weak topology. Together with Lemma 6.6 this completes the proof. □

6.1.3 Continuity of data fidelity terms

With the confirmed equi-coercivity of $(\mathcal{E}_\alpha^{m_\mu,n_\mu})_{\mu\in\mathbb{N}}$ with respect to the weak topology on $\mathcal{T}$ we now know in which sense a continuous convergence of $(\mathcal{D}_{B_{m_\mu,n_\mu}})_{\mu\in\mathbb{N}}$ should be understood in order to strive for a global Γ-limit that maintains the minimizing structure. Hence, we can verify the following proposition.

Proposition 6.7. Let (m_μ, n_μ), $\mathcal{T}$ and E be defined according to the scenario in (I), (II) or (III). Suppose that $(B_{m_\mu,n_\mu})_{\mu\in\mathbb{N}} \subset \bigcup_{\mu\in\mathbb{N}} \mathbb{R}^{m_\mu\times n_\mu}$ is a sequence of matrices whose sequence of associated operators $(EB_{m_\mu,n_\mu})_{\mu\in\mathbb{N}} \subset \mathcal{T}_2$ converges with respect to the strong topology on $\mathcal{T}_2$ toward $B \in \mathcal{T}_2$ and let $\mathcal{D}_{B_{m_\mu,n_\mu}} : \mathcal{T} \to \mathbb{R}$ be defined as in (6.8).
Then $(\mathcal{D}_{B_{m_\mu,n_\mu}})_{\mu\in\mathbb{N}}$ is continuously convergent with respect to the weak topology on $\mathcal{T}$ to $\mathcal{D}_B : \mathcal{T} \to \mathbb{R}$ with

$$\mathcal{D}_B := \frac{1}{2}||\overline{K}T - B||^2_{\mathcal{T}_2}.$$

Proof. Let $(T_\mu)_{\mu\in\mathbb{N}} \subset \mathcal{T}$ be a weakly convergent sequence with limit $T_* \in \mathcal{T}$. Then, due to Lemma 2.14 and the compactness of $\overline{K}$ it follows that $(\overline{K}T_\mu)_{\mu\in\mathbb{N}} \subset \mathcal{T}_2$ converges with respect to the strong topology toward $\overline{K}T_* \in \mathcal{T}_2$. Together with the strong convergence of $(EB_{m_\mu,n_\mu})_{\mu\in\mathbb{N}} \subset \mathcal{T}_2$ to $B \in \mathcal{T}_2$ this directly implies

$$\lim_{\mu\to\infty} \mathcal{D}_{B_{m_\mu,n_\mu}}(T_\mu) = \lim_{\mu\to\infty} \frac{1}{2}||\overline{K}T_\mu - EB_{m_\mu,n_\mu}||^2_{\mathcal{T}_2} = \frac{1}{2}||\overline{K}T_* - B||^2_{\mathcal{T}_2} = \mathcal{D}_B(T_*).$$

□

As already insinuated, this proposition together with the statement in Proposition 2.23 suggests to neglect the data fidelity terms when considering the Γ-convergence of $(\mathcal{E}_\alpha^{m_\mu,n_\mu})_{\mu\in\mathbb{N}}$. As soon as one can find a Γ-limit for the sequence of regularizers $(\mathcal{R}^{m_\mu,n_\mu})_{\mu\in\mathbb{N}}$, this, combined with $\mathcal{D}_B$, represents exactly the Γ-limit of the sequence of interest. Therefore, we want to dedicate the subsequent section exclusively to the detailed determination of the Γ-limits of regularizing sequences that involve the norms which were introduced previously in this thesis.

6.2 Γ-convergence of regularizing norms

In the previous section we showed that the sequence of functionals $(\mathcal{E}_\alpha^{m_\mu,n_\mu})_{\mu\in\mathbb{N}}$ defined in (6.7) is equi-coercive with respect to the weak topology. Beyond that, referring to the same topology, we saw that its corresponding sequence of data fidelity terms $(\mathcal{D}_{B_{m_\mu,n_\mu}})_{\mu\in\mathbb{N}}$ is continuously convergent. Keeping the statements in Theorem 2.20 and Proposition 2.23 in mind we are now interested in completing the asymptotic analysis of $(\mathcal{E}_\alpha^{m_\mu,n_\mu})_{\mu\in\mathbb{N}}$ by discussing the Γ-convergence of the sequence of involved regularizers.
Within the following subsections we therefore want to address the determination of the Γ-limits of mixed norms, the nuclear norm and the generalized nuclear norm. In order to eventually profit from the previously mentioned results these limits should all be ascertained with respect to the weak topology on $\mathcal{T}$. For the sake of comprehensibility, in doing so, we want to renounce working with placeholders $\mathcal{T}$ and $\mathfrak{T}$ to treat all three continuity approaches at once. Instead, we want to contemplate the spaces $\mathcal{I}^{q,p}(\mathbb{R}^n, \Sigma)$, $\mathcal{I}^{q,p}(\Omega, \mathbb{R}^m)$ and $\mathcal{I}^{q,p}(\Omega, \Sigma)$ separately and, in the case of upcoming similar arguments, only refer to the more extensive explanation.

6.2.1 Mixed norms

We start our analysis with the consideration of discrete regularizers that coincide with weighted mixed norms for values $p, q > 1$. Harking back to the formulation in (6.1) this means that we are interested in the case where $\mathcal{R} : \mathbb{R}^{m\times n} \to \overline{\mathbb{R}}$ is defined via

$$\mathcal{R}(A) = ||A||_{\omega;p,q} \quad \text{for } p, q > 1,\ \omega \in \mathbb{R}^{m\times n}. \tag{6.11}$$

This choice obviously fulfills the condition in (6.2) such that the results in Section 6.1 apply. Being first of all interested in the space-continuous but time-discrete continuity scenario, according to (6.9), the definition in (6.11) implies that we need to consider the asymptotic behavior of the sequence of semi-discrete functionals $\mathcal{R}^{m,n}_{p,q} : \mathcal{I}^{q,p}(\mathbb{R}^n, \Sigma) \to \overline{\mathbb{R}}$ which are defined by

$$\mathcal{R}^{m,n}_{p,q}(T) := \begin{cases} \left(\sum_{i=1}^m \left(\sum_{j=1}^n \omega_{ij}|A_{ij}|^q\right)^{\frac{p}{q}}\right)^{\frac{1}{p}}, & \text{if } T = EA \text{ for some } A \in \mathbb{R}^{m\times n} \\ \infty, & \text{else.} \end{cases} \tag{6.12}$$

For $m \to \infty$ this results in the following theorem.

Theorem 6.8. Let $p, q > 1$. Suppose that $(\mathcal{S}_m)_{m\in\mathbb{N}}$ is a sequence of partitions of Σ with the following property:

$$(\#) \qquad \max_{i\in\{1,\dots,m\}} \operatorname{diam}(\Sigma_i^m) \xrightarrow{m\to\infty} 0.$$

Let furthermore $\omega_{ij} := |\Sigma_i^m|^{\frac{q}{p}}$ for all $i \in \{1, ..., m\}$, $j \in \{1, ..., n\}$. Then $(\mathcal{R}^{m,n}_{p,q})_{m\in\mathbb{N}}$ Γ-converges with respect to the weak operator topology on $\mathcal{I}^{q,p}(\mathbb{R}^n, \Sigma)$ for $m \to \infty$ to $\mathcal{R}^{\infty,n}_{p,q} : \mathcal{I}^{q,p}(\mathbb{R}^n, \Sigma) \to \mathbb{R}$ with

$$\mathcal{R}^{\infty,n}_{p,q}(T) := \left(\int_\Sigma \left(\sum_{j=1}^n |t_j(s)|^q\right)^{\frac{p}{q}} \mathrm{d}s\right)^{\frac{1}{p}}.$$

Prior to proving this theorem we want to draw attention to a property that applies to all discretely representable integral operators.

Remark 6.9. We have already seen in Remark 6.2 that for a matrix $A \in \mathbb{R}^{m\times n}$ and $t = \left(\sum_{i=1}^m A_{ij}\chi_{\Sigma_i^m}\right)_{j=1,\dots,n} \in L^p(\Sigma; \mathbb{R}^n)$ the identity

$$||A||_{\omega;p,q} = \left(\int_\Sigma \left(\sum_{j=1}^n |t_j(s)|^q\right)^{\frac{p}{q}} \mathrm{d}s\right)^{\frac{1}{p}}$$

holds true as soon as $\omega_{ij} := |\Sigma_i^m|^{\frac{q}{p}}$. Applied to the setting in Theorem 6.8 this implies that the evaluations of $\mathcal{R}^{m,n}_{p,q}$ and $\mathcal{R}^{\infty,n}_{p,q}$ at any discretely representable integral operator $T = EA$

coincide.

This observation now enables us to prove Theorem 6.8.

Proof of Theorem 6.8.

(1) *Lim inf inequality:*
Let $(T_m)_{m\in\mathbb{N}} \subset \mathcal{I}^{q,p}(\mathbb{R}^n, \Sigma)$ be a convergent sequence with respect to the weak operator topology, i.e. there exists some T_* such that $T_m y \rightharpoonup T_* y\ \forall y \in \mathbb{R}^n$ for $m \to \infty$.
According to Lemma 6.6 this implicitly defines the weakly convergent sequence $(t_m)_{m\in\mathbb{N}}$ of elements in $L^p(\Sigma; \mathbb{R}^n)$ whose limit t_* corresponds to T_*. Consequently, $T_* \in \mathcal{I}^{q,p}(\mathbb{R}^n, \Sigma)$ holds and for $\liminf_{m\to\infty} \mathcal{R}^{m,n}_{p,q}(T_m) = \infty$ the inequality is immediately true since

$$\mathcal{R}^{\infty,n}_{p,q}(T_*) < \infty = \liminf_{m\to\infty} \mathcal{R}^{m,n}_{p,q}(T_m).$$

Hence, let $\liminf_{m\to\infty} \mathcal{R}^{m,n}_{p,q}(T_m) < \infty$. Then there exists a subsequence $(T_{m_\phi})_{\phi\in\mathbb{N}} \subseteq (T_m)_{m\in\mathbb{N}}$ such that

$$\infty > \liminf_{m\to\infty} \mathcal{R}^{m,n}_{p,q}(T_m) = \lim_{\phi\to\infty} \mathcal{R}^{m_\phi,n}_{p,q}(T_{m_\phi})$$

which implies

$$\sup_{\phi\in\mathbb{N}} \mathcal{R}^{m_\phi,n}_{p,q}(T_{m_\phi}) < \infty$$

and thus the boundedness of the associated subsequence $(t_{m_\phi})_{\phi\in\mathbb{N}} \subseteq (t_m)_{m\in\mathbb{N}}$. Taking into account the reflexivity of $L^p(\Sigma; \mathbb{R}^n)$ (cf. Remark 2.2), with Banach-Alaoglu we can therefore find a subsubsequence $(t_{m_{\phi_\psi}})_{\psi\in\mathbb{N}} \subseteq (t_{m_\phi})_{\phi\in\mathbb{N}} \subseteq (t_m)_{m\in\mathbb{N}}$ and a $\widetilde{t} \in L^p(\Sigma; \mathbb{R}^n)$ such that $t_{m_{\phi_\psi}} \rightharpoonup \widetilde{t}$ for $\psi \to \infty$. This, again due to Lemma 6.6, characterizes the weakly convergent subsubsequence $(T_{m_{\phi_\psi}})_{\psi\in\mathbb{N}} \subseteq (T_{m_\phi})_{\phi\in\mathbb{N}} \subseteq (T_m)_{m\in\mathbb{N}}$ with limit $\widetilde{T} \in \mathcal{I}^{q,p}(\mathbb{R}^n, \Sigma)$ which thanks to the uniqueness of the weak convergence coincides with T_*.
Using the correspondence between $(T_{m_{\phi_\psi}})_{\psi\in\mathbb{N}}$ and $(t_{m_{\phi_\psi}})_{\psi\in\mathbb{N}}$ as well as between their respective limits together with the lower semicontinuity of the norm on $L^p(\Sigma; \mathbb{R}^n)$ (cf. Remark 2.10) eventually leads to

$$\begin{aligned}
\liminf_{m\to\infty} \mathcal{R}^{m,n}_{p,q}(T_m) &= \lim_{\phi\to\infty} \mathcal{R}^{m_\phi,n}_{p,q}(T_{m_\phi}) \\
&= \lim_{\psi\to\infty} \mathcal{R}^{m_{\phi_\psi},n}_{p,q}(T_{m_{\phi_\psi}}) \\
&= \liminf_{\psi\to\infty} \mathcal{R}^{m_{\phi_\psi},n}_{p,q}(T_{m_{\phi_\psi}}) \\
&= \liminf_{\psi\to\infty} \mathcal{R}^{\infty,n}_{p,q}(T_{m_{\phi_\psi}}) \\
&\geq \mathcal{R}^{\infty,n}_{p,q}(T_*).
\end{aligned}$$

Notice that in the fourth equality we used the observation in Remark 6.9.

(2) *Existence of a recovery sequence:*
Let $\mathcal{I}_C^{q,p}(\mathbb{R}^n,\Sigma)$ be the set of all integral operators which are induced by $t \in C_c^\infty(\Sigma;\mathbb{R}^n)$, i.e.

$$\mathcal{I}_C^{q,p}(\mathbb{R}^n,\Sigma) := \left\{ T : \mathbb{R}^n \to L^p(\Sigma) \,\middle|\, (Ty)(s) = \sum_{j=1}^n y_j t_j(s) \text{ for } t \in C_c^\infty(\Sigma;\mathbb{R}^n) \right\}.$$

Then $\mathcal{I}_C^{q,p}(\mathbb{R}^n,\Sigma)$ lies dense in $\mathcal{I}^{q,p}(\mathbb{R}^n,\Sigma)$ since $C_c^\infty(\Sigma;\mathbb{R}^n)$ lies dense in $L^p(\Sigma;\mathbb{R}^n)$ (cf. Corollary 2.4) and it suffices to show the existence of a recovery sequence for all $T \in \mathcal{I}_C^{q,p}(\mathbb{R}^n,\Sigma)$ (cf. Remark 2.21).
Thus, let $T \in \mathcal{I}_C^{q,p}(\mathbb{R}^n,\Sigma)$ and $t \in C_c^\infty(\Sigma;\mathbb{R}^n)$ be its corresponding kernel characterizing element. We define a sequence of integral operators $(T_m)_{m\in\mathbb{N}} \subset \mathcal{I}^{q,p}(\mathbb{R}^n,\Sigma)$ via the sequence of its inducing elements $(t_m)_{m\in\mathbb{N}} \subset L^p(\Sigma;\mathbb{R}^n)$ which we characterize by

$$(t_m)_j(s) := \sum_{i=1}^m \frac{\int_{\Sigma_i^m} t_j(\tilde{s})\,\mathrm{d}\tilde{s}}{|\Sigma_i^m|} \chi_{\Sigma_i^m}(s) \quad \forall j \in \{1,\dots,n\}, s \in \Sigma.$$

Due to the continuity of the t_j and property (#) we can then find for all $\varepsilon_j > 0$ some $M_j \in \mathbb{N}$ such that for $s \in \Sigma_{i'}^m$, $i' \in \{1,\dots,m\}$ and all $m \geq M_j$ the following holds:

$$\begin{aligned} |(t_m)_j(s) - t_j(s)| &= \left| \frac{\int_{\Sigma_{i'}^m} t_j(\tilde{s})\,\mathrm{d}\tilde{s}}{|\Sigma_{i'}^m|} - t_j(s) \right| \\ &= \frac{1}{|\Sigma_{i'}^m|} \left| \int_{\Sigma_{i'}^m} t_j(\tilde{s}) - t_j(s)\,\mathrm{d}\tilde{s} \right| \\ &\leq \sup_{\tilde{s}\in\Sigma_{i'}^m} |t_j(\tilde{s}) - t_j(s)| \\ &\leq \varepsilon_j. \end{aligned} \tag{6.13}$$

The combination of this result for varying $j \in \{1,\dots,n\}$ then yields that for every $\varepsilon > 0$ there also exists some $M \in \mathbb{N}$ which guarantees $|t_m - t| \leq \varepsilon$ for all $m \geq M$. Hence, the above constructed sequence of elements in $L^p(\Sigma;\mathbb{R}^n)$ converges (uniformly) toward t implying also the convergence of $(T_m)_{m\in\mathbb{N}}$ toward T. Together with the continuity of the single components of $\mathcal{R}_{p,q}^{\infty,n}$ and the coincidence of $\mathcal{R}_{p,q}^{m,n}(T_m)$ and $\mathcal{R}_{p,q}^{\infty,n}(T_m)$ for all $m \in \mathbb{N}$ this ultimately leads to

$$\lim_{m\to\infty} \mathcal{R}_{p,q}^{m,n}(T_m) = \lim_{m\to\infty} \mathcal{R}_{p,q}^{\infty,n}(T_m) = \mathcal{R}_{p,q}^{\infty,n}(T).$$

□

When considering the contrary continuity approach in which one strives for a time-continuous but space-discrete scenario, i.e. when contemplating $\mathcal{I}^{q,p}(\Omega,\mathbb{R}^m)$, the discrete choice in (6.11) asks for the asymptotic study of the sequence of functionals $\mathcal{R}_{p,q}^{m,n} : \mathcal{I}^{q,p}(\Omega,\mathbb{R}^m) \to \overline{\mathbb{R}}$ which are also characterized through the expression in (6.12). In this setting then the following convergence behavior can be observed.

Theorem 6.10. Let p,q >1. Suppose that $(\mathcal{O}_n)_{n\in\mathbb{N}}$ is a sequence of partitions of Ω with the following property:

$$(\#) \qquad \max_{j\in\{1,\ldots,n\}} \operatorname{diam}(\Omega_j^n) \xrightarrow{n\to\infty} 0.$$

Let furthermore $\omega_{ij} := |\Omega_j^n|$ for all $i \in \{1,\ldots,m\}$, $j \in \{1,\ldots,n\}$. Then $(\mathcal{R}_{p,q}^{m,n})_{n\in\mathbb{N}}$ Γ-converges with respect to the weak operator topology on $\mathcal{I}^{q,p}(\Omega,\mathbb{R}^m)$ for $n\to\infty$ to $\mathcal{R}_{p,q}^{m,\infty}$: $\mathcal{I}^{q,p}(\Omega,\mathbb{R}^m) \to \mathbb{R}$ with

$$\mathcal{R}_{p,q}^{m,\infty}(T) := \left(\sum_{i=1}^{m} ||t_i||_{L^q(\Omega)}^p\right)^{\frac{1}{p}}.$$

Similar to Remark 6.9 we can deduce from Remark 6.2 that $\mathcal{R}_{p,q}^{m,n}$ and $\mathcal{R}_{p,q}^{m,\infty}$ coincide for discretely representable arguments if $\omega_{ij} := |\Omega_j^n|$. Since also here due to Lemma 6.6 the weak convergences of $(T_n)_{n\in\mathbb{N}} \subset \mathcal{I}^{q,p}(\Omega,\mathbb{R}^m)$ and $(t_n)_{n\in\mathbb{N}} \subset L^q(\Omega;\mathbb{R}^m)$ are equivalent, this theorem can be proven following the same line of argument of the proof of Theorem 6.8 which is why we want to omit further details.

Covering the last approach, which simultaneously provides continuity in space and time, we interpret the expression in (6.12) as the definition of a functional that maps from $\mathcal{I}^{q,p}(\Omega,\Sigma)$ to $\overline{\mathbb{R}}$. In this scenario we are interested in the behavior of $\mathcal{R}_{p,q}^{m,n}$ as soon as both parameters, m and n, tend toward infinity. In order to avoid notational confusion while pursuing this interest we want to revisit subscripting the pair (m,n). By considering (m_μ, n_μ) with $m_\mu, n_\mu \to \infty$ as $\mu\to\infty$ we are then able to observe the limit behavior dependent on a single parameter. Unsurprisingly, the according analysis results in the combination of the previous two assertions.

Theorem 6.11. Let $p,q>1$. Suppose that $(\mathcal{S}_{m_\mu})_{\mu\in\mathbb{N}}$, $(\mathcal{O}_{n_\mu})_{\mu\in\mathbb{N}}$ are sequences of partitions of Σ and Ω with the following properties:

$$(\#) \qquad \max_{i\in\{1,\ldots,m_\mu\}} \operatorname{diam}(\Sigma_i^{m_\mu}) \xrightarrow{\mu\to\infty} 0,$$

$$(\#\#) \qquad \max_{j\in\{1,\ldots,n_\mu\}} \operatorname{diam}(\Omega_j^{n_\mu}) \xrightarrow{\mu\to\infty} 0.$$

Let furthermore $\omega_{ij} := |\Omega_j^{n_\mu}||\Sigma_i^{m_\mu}|^{\frac{q}{p}}$ for all $i\in\{1,...,m_\mu\}$, $j\in\{1,...,n_\mu\}$. Then $(\mathcal{R}_{p,q}^{m_\mu,n_\mu})_{\mu\in\mathbb{N}}$ Γ-converges with respect to the weak operator topology on $\mathcal{I}^{q,p}(\Omega,\Sigma)$ for $\mu\to\infty$ to $\mathcal{R}_{p,q}^{\infty}$: $\mathcal{I}^{q,p}(\Omega,\Sigma)\to\mathbb{R}$ with

$$\mathcal{R}_{p,q}^{\infty}(T) := ||t||_{L^{p,q}(\Sigma\times\Omega)}.$$

The similarity to the assertions in Theorem 6.8 and Theorem 6.10 also allows us to again proceed equally when it comes to proving. The statements in Remark 6.2 and Lemma 6.6 can be applied likewise. Incorporating, in addition to that, the lower semicontinuity of the norm on $L^{p,q}(\Sigma \times \Omega)$ regarding the weak topology and the denseness of $C_c^\infty(\Sigma \times \Omega)$ in $L^{p,q}(\Sigma \times \Omega)$ (cf. Remark 2.10, Lemma 2.11), we can eventually pursue the exact same strategy as in the proof of Theorem 6.8.

With the results in Theorems 6.8, 6.10 and 6.11 we therefore found limits that, depending on the respective type of continuity approach, complete the asymptotic analysis of functionals of the form (6.7) that include regularizing mixed norms. Combined with the respective results in Propositions 6.5 and 6.7 they guarantee the minimization structure preserving Γ-convergence of $\mathcal{E}_\alpha^{m,n}$.

6.2.2 Nuclear norm

After the warm-up in the previous subsection we now want to turn toward the analysis of the asymptotic behavior of the nuclear norm. In terms of the formulation in (6.1) this means we are interested in considering problems that involve

$$\mathcal{R}(A) = ||A||_{\omega;*} \quad \text{for } \omega \in \mathbb{R}^{m\times n}.$$

For $p = q = 2$, the condition in (6.2) can also be verified in this case: Choosing $\overline{\omega} \in \mathbb{R}^{m\times n}$ to be defined by $\overline{\omega}_{ij} = \omega_{ij}^2$ for all $i \in \{1, \ldots, m\}$, $j \in \{1, \ldots n\}$ while taking advantage of the general positivity of all singular values and the representation of the Frobenius norm over the singular value decomposition we find that

$$||A||_{\overline{\omega};2,2} = ||\omega \cdot A||_{2,2} = \left(\sum_{\xi=1}^{\min(m,n)} \sigma_\xi(\omega \cdot A)^2 \right)^{\frac{1}{2}} \tag{6.14}$$

$$\leq \left(\sigma_{\max}(\omega \cdot A) \sum_{\xi=1}^{\min(m,n)} \sigma_\xi(\omega \cdot A) \right)^{\frac{1}{2}} \tag{6.15}$$

$$\leq \left(\left(\sum_{\xi=1}^{\min(m,n)} \sigma_\xi(\omega \cdot A) \right)^2 \right)^{\frac{1}{2}} = ||A||_{\omega;*}. \tag{6.16}$$

Consequently, the conclusions in Section 6.1 are applicable and studying the asymptotic behavior of the nuclear norm turns out to be the missing piece of the puzzle on our way to grasp how functionals $\mathcal{E}_\alpha^{m,n}$ defined by (6.7) and their corresponding minimizers act as soon as dimensions tend to infinity.

Concerning the analysis of the mixed norm we saw that, regardless of the contemplated continuity approach, the proofs of all Γ-convergence results were mainly following the same strategy. This was because, realizing that there is a natural bijective relation between $\mathcal{I}^{q,p}(\mathbb{R}^n, \Sigma)$ and

$$\mathcal{I}^{q,p}_n(\Omega,\Sigma) := \left\{ T \in \mathcal{I}^{q,p}(\Omega,\Sigma) \middle| (Ty)(s) = \int_\Omega \sum_{j=1}^{n} t_j(s)\chi_{\Omega_j^n}(r)y(r)\,\mathrm{d}r \text{ with } t \in L^p(\Sigma;\mathbb{R}^n) \right\} \tag{6.17}$$

for any $\Omega \subset \mathbb{R}^d$ with arbitrary partition $\mathcal{O}_n = \{\Omega_1^n, \dots, \Omega_n^n\}$, respectively between $\mathcal{I}^{q,p}(\Omega, \mathbb{R}^m)$ and

$$\mathcal{I}^{q,p}_m(\Omega,\Sigma) := \left\{ T \in \mathcal{I}^{q,p}(\Omega,\Sigma) \middle| (Ty)(s) = \int_\Omega \sum_{i=1}^{m} t_i(r)\chi_{\Sigma_i^m}(s)y(r)\,\mathrm{d}r \text{ with } t \in L^q(\Omega;\mathbb{R}^m) \right\} \tag{6.18}$$

for any $\Sigma \subset \mathbb{R}^d$ with arbitrary partition $\mathcal{S}_m = \{\Sigma_1^m, \dots, \Sigma_m^m\}$, both semi-discrete scenarios can in general be perceived as special cases of the entirely continuous case. Within this subsection we thus want to swap the order in which we are considering the various continuity approaches. Starting with the most general one - the one which aims for a continuous resolution in space *and* time - we can refer to similar proving techniques more efficiently and therefore facilitate the ensuing elaborations.
Hence, reverting to the definition in (6.9) we are first of all intrigued in understanding the behavior of the functional $\mathcal{R}_*^{m,n} : \mathcal{I}^{2,2}(\Omega,\Sigma) \to \overline{\mathbb{R}}$ characterized by

$$\mathcal{R}_*^{m,n}(T) := \begin{cases} \sum_{\xi=1}^{\min(m,n)} \sigma_\xi(\omega \cdot A), & \text{if } T = EA \text{ for some } A \in \mathbb{R}^{m\times n} \\ \infty, & \text{else} \end{cases} \tag{6.19}$$

for $\omega \in \mathbb{R}^{m\times n}$ as soon as $m, n \to \infty$. With this objective, harking back to the single variable notation (m_μ, n_μ) used in Theorem 6.11, we examine its Γ-limit for $\mu \to \infty$.

Theorem 6.12. Suppose that $(\mathcal{S}_{m_\mu})_{\mu\in\mathbb{N}}$, $(\mathcal{O}_{n_\mu})_{\mu\in\mathbb{N}}$ are sequences of partitions of Σ and Ω with the following properties:

$$(\#) \qquad \max_{i\in\{1,\dots,m_\mu\}} \operatorname{diam}(\Sigma_i^{m_\mu}) \xrightarrow{\mu\to\infty} 0,$$
$$(\#\#) \qquad \max_{j\in\{1,\dots,n_\mu\}} \operatorname{diam}(\Omega_j^{n_\mu}) \xrightarrow{\mu\to\infty} 0.$$

Let furthermore $\omega_{ij} := (|\Sigma_i^{m_\mu}||\Omega_j^{n_\mu}|)^{\frac{1}{2}}$ for all $i \in \{1,\dots,m_\mu\}$, $j \in \{1,\dots,n_\mu\}$ and $(f_\phi)_{\phi\in\mathbb{N}}$ be an orthonormal basis of $L^2(\Omega)$. Then $(\mathcal{R}_*^{m_\mu,n_\mu})_{\mu\in\mathbb{N}}$ Γ-converges with respect to the weak operator topology on $\mathcal{I}^{2,2}(\Omega,\Sigma)$ for $\mu\to\infty$ to $\mathcal{R}_*^\infty : \mathcal{I}^{2,2}(\Omega,\Sigma) \to \overline{\mathbb{R}}$ with

$$\mathcal{R}_*^\infty(T) := \sum_{\phi\in\mathbb{N}} \left\langle (T^*T)^{\frac{1}{2}} f_\phi \,,\, f_\phi \right\rangle .$$

Though the proof of this theorem is based on very similar components as the ones related to the mixed norm there are some crucial points where we have to work a little bit harder. This is why we want to give a detailed explanation on its derivation instead of just pointing to the differences. For this purpose, we start by collecting some auxiliary statements which will prove useful in the following.

Lemma 6.13. Let $T \in \mathcal{I}^{2,2}(\Omega, \Sigma)$ and $(\sigma_\xi)_{\xi\in\mathbb{N}}$ be its singular values. Then

$$\mathcal{R}_*^\infty(T) = \sum_{\xi\in\mathbb{N}} \sigma_\xi.$$

Proof. Let $T \in \mathcal{I}^{2,2}(\Omega, \Sigma)$ and $(\sigma_\xi, u_\xi, v_\xi)_{\xi\in\mathbb{N}}$ be an associated singular system. Hence, by definition $(\sigma_\xi^2, v_\xi)_{\xi\in\mathbb{N}}$ is an eigensystem of the selfadjoint operator T^*T which therefore, due to the spectral theorem, can be understood as

$$T^*T = \sum_{\xi\in\mathbb{N}} \sigma_\xi^2 (v_\xi \otimes v_\xi).$$

Defining $B : L^2(\Omega) \to L^2(\Omega)$ via

$$B := \sum_{\xi\in\mathbb{N}} \sigma_\xi (v_\xi \otimes v_\xi)$$

we consequently compute that for every $y \in L^2(\Omega)$

$$B(By) = \sum_{\xi\in\mathbb{N}} \sigma_\xi \left\langle \sum_{\eta\in\mathbb{N}} \sigma_\eta \langle y, v_\eta\rangle v_\eta \, , \, v_\xi \right\rangle v_\xi = \sum_{\xi\in\mathbb{N}} \sigma_\xi^2 \langle y, v_\xi\rangle v_\xi = T^*Ty$$

applies which indicates the equality of B and $(T^*T)^{\frac{1}{2}}$. Following the definition of $\mathcal{R}_*^\infty$ and making use of Parseval's identity then eventually yields for any orthonormal basis $(f_\phi)_{\phi\in\mathbb{N}}$ of $L^2(\Omega)$

$$\mathcal{R}_*^\infty(T) = \sum_{\phi\in\mathbb{N}} \langle Bf_\phi, f_\phi\rangle = \sum_{\phi\in\mathbb{N}} \left\langle \sum_{\xi\in\mathbb{N}} \sigma_\xi \langle f_\phi, v_\xi\rangle v_\xi \, , \, f_\phi \right\rangle = \sum_{\xi\in\mathbb{N}} \sigma_\xi \sum_{\phi\in\mathbb{N}} |\langle f_\phi, v_\xi\rangle|^2 = \sum_{\xi\in\mathbb{N}} \sigma_\xi.$$

□

Note that with the help of this lemma we furthermore gain information on the relation between $\mathcal{R}_*^{m,n}$ and $\mathcal{R}_*^\infty$ for discretely representable integral operators.

Remark 6.14. For $\omega_{ij} := (|\Sigma_i||\Omega_j|)^{\frac{1}{2}}$ and any $T = EA$ we can calculate that

$$\begin{aligned}(T^*Ty)(r) &= \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}\chi_{\Omega_j^n}(r)\int_{\Sigma_i^m}(Ty)(s)\,\mathrm{d}s\\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}\chi_{\Omega_j^n}(r)\int_{\Sigma_i^m}\sum_{i'=1}^{m}\sum_{j'=1}^{n} A_{i'j'}\chi_{\Sigma_{i'}^m}(s)\left(\int_{\Omega_{j'}^n} y(\bar r)\,\mathrm{d}\bar r\right)\mathrm{d}s\\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}|\Sigma_i^m|\chi_{\Omega_j^n}(r)\left(\sum_{j'=1}^{n} A_{ij'}\left(\int_{\Omega_{j'}^n} y(\bar r)\,\mathrm{d}\bar r\right)\right)\end{aligned}$$

holds for any $y \in L^2(\Omega)$, $r \in \Omega$, such that the eigenvalue problem for T^*T, which yields the singular values of T, can at first be described by

$$\sum_{j=1}^{n}\left[\sum_{i=1}^{m} A_{ij}|\Sigma_i^m|\left(\sum_{j'=1}^{n} A_{ij'}\left(\int_{\Omega_{j'}^n} v(\bar r)\,\mathrm{d}\bar r\right)\right)\right]\chi_{\Omega_j^n}(r) = \lambda v(r) \quad \forall r \in \Omega$$

for an eigenvalue $\lambda \in \mathbb{R}$ and an eigenfunction $v \in L^2(\Omega)$. Reinserting the information thus obtained about the representation of the eigenfunction, namely that

$$v(r) = \sum_{j=1}^{n} \frac{c_j}{|\Omega_j^n|^{\frac{1}{2}}}\chi_{\Omega_j^n}(r) \quad \forall r \in \Omega$$

for some $c \in \mathbb{R}^n$, as well as considering the linear independence of the $\chi_{\Omega_j^n}$ returns

$$\begin{aligned}& T^*Tv = \lambda v\\ \Leftrightarrow\quad & \sum_{j=1}^{n}\left[\sum_{i=1}^{m} A_{ij}\,|\Sigma_i|\left(\sum_{j'=1}^{n} A_{ij'}\,c_{j'}\,|\Omega_{j'}|^{\frac{1}{2}}\right)\right]\chi_{\Omega_j}(r) = \lambda\sum_{j=1}^{n}\frac{c_j}{|\Omega_j|^{\frac{1}{2}}}\chi_{\Omega_j}(r) \quad \forall r \in \Omega\\ \Leftrightarrow\quad & \sum_{i=1}^{m}(|\Sigma_i^m||\Omega_j^n|)^{\frac{1}{2}}A_{ij}\left(\sum_{j'=1}^{n}(|\Sigma_i^m||\Omega_{j'}^n|)^{\frac{1}{2}}A_{ij'}c_{j'}\right) = \lambda c_j \quad \forall j = 1,\dots,n\\ \Leftrightarrow\quad & (\omega\cdot A)^T(\omega\cdot A)c = \lambda c.\end{aligned}$$

Thus, the singular values of T match exactly with those of the matrix $\omega \cdot A$ which together with Lemma 6.13 guarantees the equality of $\mathcal{R}_*^{m,n}(T)$ and $\mathcal{R}_*^{\infty}(T)$.

Another useful tool which helps proving Theorem 6.12 is the lower semicontinuity of $\mathcal{R}_*^{\infty}$. Its derivation makes use of the following lemma.

Lemma 6.15. Let $\widehat{\mathcal{I}}^{2,2}(\Omega,\Sigma)$ be the set of all operators in $\mathcal{I}^{2,2}(\Omega,\Sigma)$ with finite number of nonzero singular values, i.e.

$$\widehat{\mathcal{I}}^{2,2}(\Omega,\Sigma) := \{T \in \mathcal{I}^{2,2}(\Omega,\Sigma) \,|\, \exists \Xi \in \mathbb{N} : \sigma_\xi(T) = 0 \,\forall \xi > \Xi\}.$$

Suppose that $\langle\cdot,\cdot\rangle$ denotes the inner product on $\mathcal{I}^{2,2}(\Omega,\Sigma)$ defined in Remark 6.4. Then, for any $T \in \mathcal{I}^{2,2}(\Omega,\Sigma)$

$$\mathcal{R}_*^\infty(T) = \sup_{\substack{L\in\widehat{\mathcal{I}}^{2,2}(\Omega,\Sigma)\\ \sigma_{\max}(L)\le 1}} \langle L, T\rangle$$

applies.

Proof. Let $T \in \mathcal{I}^{2,2}(\Omega,\Sigma)$ and $(\sigma_\xi, u_\xi, v_\xi)_{\xi\in\mathbb{N}}$ be an associated singular system. In order to prove the assertion in the following we want to exhibit:

(i) $\sup_{\substack{L\in\widehat{\mathcal{I}}^{2,2}(\Omega,\Sigma)\\ \sigma_{\max}(L)\le 1}} \langle L, T\rangle \ge \mathcal{R}_*^\infty(T)$,

(ii) $\sup_{\substack{L\in\widehat{\mathcal{I}}^{2,2}(\Omega,\Sigma)\\ \sigma_{\max}(L)\le 1}} \langle L, T\rangle \le \mathcal{R}_*^\infty(T)$.

For the first conjecture we define the integral operator $\overline{L} \in \mathcal{I}^{2,2}(\Omega,\Sigma)$ via its kernel

$$\overline{\ell}(s,r) := \sum_{\xi\in\mathbb{N}} u_\xi(s)\, v_\xi(r) \quad \forall s \in \Sigma,\, \forall r \in \Omega.$$

By computing the eigenvalues of $\overline{L}^*\overline{L}$ we can verify that all singular values of $\overline{L}$ equal 1, such that $\sigma_{\max}(\overline{L}) = 1$ applies. Now utilizing the representation of T through its singular value decomposition, i.e. realizing that its characterizing element $t \in L^2(\Sigma\times\Omega)$ is of the form

$$t(s,r) = \sum_{\xi\in\mathbb{N}} \sigma_\xi u_\xi(s) v_\xi(r) \quad \forall s \in \Sigma, r \in \Omega$$

(cf. Lemma 2.17), then leads to

$$\begin{aligned}
\langle \overline{L}, T\rangle &= \int_\Sigma \int_\Omega \overline{\ell}(s,r) t(s,r)\,\mathrm{d}r\,\mathrm{d}s \\
&= \sum_{\xi\in\mathbb{N}} \sum_{\eta\in\mathbb{N}} \sigma_\eta \int_\Sigma u_\xi(s) u_\eta(s)\,\mathrm{d}s \int_\Omega v_\xi(r) v_\eta(r)\,\mathrm{d}r \\
&= \sum_{\xi\in\mathbb{N}} \sigma_\xi.
\end{aligned}$$

In fact we can even deduce likewise that for a truncated version $\overline{L}_\Xi \in \widehat{\mathcal{I}}^{2,2}(\Omega,\Sigma)$ of $\overline{L}$, which for some $\Xi \in \mathbb{N}$ is defined by the kernel

$$\overline{\ell}_{\Xi}(s,r) := \sum_{\xi=1}^{\Xi} u_{\xi}(s) v_{\xi}(r) \quad \forall s \in \Sigma, r \in \Omega$$

and therefore shares the maximum singular value of $\overline{L}$,

$$\langle \overline{L}_{\Xi}, T \rangle = \sum_{\xi=1}^{\Xi} \sigma_{\xi}$$

holds true. Combining these identities and incorporating the result in Lemma 6.13 we thus obtain that

$$\mathcal{R}_*^{\infty}(T) = \langle \overline{L}_{\Xi}, T \rangle + \sum_{\xi=\Xi+1}^{\infty} \sigma_{\xi} \leq \sup_{\substack{L \in \widehat{\mathcal{I}}^{2,2}(\Omega,\Sigma) \\ \sigma_{\max}(L) \leq 1}} \langle L, T \rangle + \sum_{\xi=\Xi+1}^{\infty} \sigma_{\xi}. \tag{6.20}$$

Since $\Xi \in \mathbb{N}$ was arbitrary and the latter summand in (6.20) is tending to zero as Ξ tends to infinity this eventually implies

$$\sup_{\substack{L \in \widehat{\mathcal{I}}^{2,2}(\Omega,\Sigma) \\ \sigma_{\max}(L) \leq 1}} \langle L, T \rangle \geq \mathcal{R}_*^{\infty}(T).$$

For the second conjecture we want to contemplate an arbitrary $L \in \mathcal{I}^{2,2}(\Omega, \Sigma)$ whose maximum singular value $\sigma_{\max}(L)$ does not exceed 1. Once again representing T and L through their singular systems $(\sigma_{\xi}, u_{\xi}, v_{\xi})_{\xi \in \mathbb{N}}$ and $(\overline{\sigma}_{\eta}, \overline{u}_{\eta}, \overline{v}_{\eta})_{\eta \in \mathbb{N}}$ and applying the inequalities of Cauchy-Schwarz and Bessel we observe that this setting implies

$$\begin{aligned}
\langle L, T \rangle &= \int_{\Sigma} \int_{\Omega} \left(\sum_{\xi \in \mathbb{N}} \sigma_{\xi} u_{\xi}(s) v_{\xi}(r) \right) \left(\sum_{\eta \in \mathbb{N}} \overline{\sigma}_{\eta} \overline{u}_{\eta}(s) \overline{v}_{\eta}(r) \right) \mathrm{d}r \, \mathrm{d}s \\
&\leq \sum_{\xi \in \mathbb{N}} \sum_{\eta \in \mathbb{N}} \sigma_{\xi} \langle u_{\xi}, \overline{u}_{\eta} \rangle \langle v_{\xi}, \overline{v}_{\eta} \rangle \\
&\leq \sum_{\xi \in \mathbb{N}} \sigma_{\xi} \left(\sum_{\eta \in \mathbb{N}} \langle u_{\xi}, \overline{u}_{\eta} \rangle^2 \right)^{\frac{1}{2}} \left(\sum_{\eta \in \mathbb{N}} \langle v_{\xi}, \overline{v}_{\eta} \rangle^2 \right)^{\frac{1}{2}} \\
&\leq \sum_{\xi \in \mathbb{N}} \sigma_{\xi} \, ||u_{\xi}||_{L^2(\Sigma)} \, ||v_{\xi}||_{L^2(\Omega)} \\
&= \sum_{\xi \in \mathbb{N}} \sigma_{\xi}.
\end{aligned}$$

Therefore, due to $\widehat{\mathcal{I}}^{2,2}(\Omega,\Sigma) \subset \mathcal{I}^{2,2}(\Omega,\Sigma)$ and the result in Lemma 6.13 we can guarantee that

$$\sup_{\substack{L \in \widehat{\mathcal{I}}^{2,2}(\Omega,\Sigma) \\ \sigma_{\max}(L) \le 1}} \langle L, T\rangle \le \sup_{\substack{L \in \mathcal{I}^{2,2}(\Omega,\Sigma) \\ \sigma_{\max}(L) \le 1}} \langle L, T\rangle \le \sum_{\xi \in \mathbb{N}} \sigma_\xi = \mathcal{R}_*^\infty(T).$$

□

With this statement we can consequently derive the lower semicontinuity of $\mathcal{R}_*^\infty$.

Lemma 6.16. $\mathcal{R}_*^\infty$ is lower semi-continuous with respect to the weak topology.

Proof. Let $L \in \widehat{\mathcal{I}}^{2,2}(\Omega,\Sigma)$ with corresponding singular system $(\sigma_\eta, u_\eta, v_\eta)_{\eta \in \{1,\dots,H\}}$ and maximum singular value that does not exceed 1. Further let $(T_\gamma)_{\gamma \in \mathbb{N}} \subset \mathcal{I}^{2,2}(\Omega,\Sigma)$ be a weakly convergent sequence of integral operators with limit T_* and respective associated characterizing elements $((t_\gamma)_\xi)_{\xi \in \mathbb{N}}$ and t_* in $L^2(\Sigma \times \Omega)$. This convergence is then equivalent to the validity of

$$\int_\Sigma x(s)\,(T_* y)(s)\,\mathrm{d}s = \lim_{\gamma \to \infty} \int_\Sigma x(s)\,(T_\gamma y)(s)\,\mathrm{d}s \quad \forall x \in L^2(\Sigma),\, \forall y \in L^2(\Omega)$$

whereby via the representation of the characterizing element of L through its singular system (cf. Lemma 2.17) it holds that

$$\begin{aligned}
\langle L, T_*\rangle &= \int_\Sigma \int_\Omega \left(\sum_{\eta=1}^{H} \sigma_\eta u_\eta(s) v_\eta(r) \right) t_*(s,r)\,\mathrm{d}r\,\mathrm{d}s \\
&= \sum_{\eta=1}^{H} \sigma_\eta \int_\Sigma u_\eta(s)(T_* v_\eta)(s)\,\mathrm{d}s \\
&= \sum_{\eta=1}^{H} \sigma_\eta \lim_{\gamma \to \infty} \int_\Sigma u_\eta(s)(T_\gamma v_\eta)(s)\,\mathrm{d}s \\
&= \lim_{\gamma \to \infty} \int_\Sigma \int_\Omega \sum_{\eta=1}^{H} \sigma_\eta u_\eta(s) v_\eta(r)\, t_\gamma(s,r)\,\mathrm{d}r\,\mathrm{d}s \\
&= \lim_{\gamma \to \infty} \langle L, T_\gamma\rangle.
\end{aligned}$$

Additionally applying Lemma 6.15 which yields for every $\gamma \in \mathbb{N}$

$$\langle L, T_\gamma\rangle \le \sup_{\substack{\overline{L} \in \widehat{\mathcal{I}}^{2,2}(\Omega,\Sigma) \\ \sigma_{\max}(\overline{L}) \le 1}} \langle \overline{L}, T_\gamma\rangle = \mathcal{R}_*^\infty(T_\gamma)$$

then leads to

$$\langle L, T_*\rangle = \liminf_{\gamma \to \infty} \langle L, T_\gamma\rangle \le \liminf_{\gamma \to \infty} \mathcal{R}_*^\infty(T_\gamma).$$

Modifying this last inequality by taking the supremum over all $L \in \widehat{\mathcal{I}}^{2,2}(\Omega,\Sigma)$ that fulfill

$\sigma_{\max}(L) \leq 1$ on the left hand side together with the identity in Lemma 6.15 eventually proves the assertion. □

Before proving Theorem 6.12 we now need to verify one last statement.

Lemma 6.17. For all $T \in \mathcal{I}^{2,2}(\Omega, \Sigma)$ it holds that

$$\mathcal{R}_*^\infty(T) = \sup \sum_{\phi\in\mathbb{N}} |\langle Tf_\phi, e_\phi\rangle|$$

where the supremum is taken over all orthonormal systems $(e_\phi)_{\phi\in\mathbb{N}}$, $(f_\phi)_{\phi\in\mathbb{N}}$ in $L^2(\Sigma)$, respectively $L^2(\Omega)$.

Proof. Let $T \in \mathcal{I}^{2,2}(\Omega, \Sigma)$ and $(\sigma_\xi, u_\xi, v_\xi)_{\xi\in\mathbb{N}}$ be an associated singular system. Thus, $(u_\xi)_{\xi\in\mathbb{N}}$ and $(v_\xi)_{\xi\in\mathbb{N}}$ are orthonormal systems in $L^2(\Sigma)$, respectively $L^2(\Omega)$, and we can deduce

$$\sup \sum_{\phi\in\mathbb{N}} |\langle Tf_\phi, e_\phi\rangle| \geq \sum_{\xi\in\mathbb{N}} |\langle Tv_\xi, u_\xi\rangle| = \sum_{\xi\in\mathbb{N}} \sigma_\xi \,||u_\xi||^2 = \mathcal{R}_*^\infty(T).$$

On the other hand, when exploiting the singular system based representation of T as well as Young's and Bessel's inequalities, we can see that for arbitrary orthonormal systems $(\bar{e}_\phi)_{\phi\in\mathbb{N}} \subset L^2(\Sigma)$ and $(\bar{f}_\phi)_{\phi\in\mathbb{N}} \subset L^2(\Omega)$

$$\begin{aligned}
\sum_{\phi\in\mathbb{N}} |\langle T\bar{f}_\phi, \bar{e}_\phi\rangle| &= \sum_{\phi\in\mathbb{N}} \left| \left\langle \sum_{\xi\in\mathbb{N}} \sigma_\xi \langle v_\xi, \bar{f}_\phi\rangle u_\xi, \bar{e}_\phi \right\rangle \right| \\
&\leq \sum_{\phi\in\mathbb{N}} \sum_{\xi\in\mathbb{N}} \sigma_\xi \left| \langle v_\xi, \bar{f}_\phi\rangle \langle u_\xi, \bar{e}_\phi\rangle \right| \\
&\leq \frac{1}{2} \sum_{\phi\in\mathbb{N}} \sum_{\xi\in\mathbb{N}} \sigma_\xi \left(|\langle v_\xi, \bar{f}_\phi\rangle|^2 + |\langle u_\xi, \bar{e}_\phi\rangle|^2 \right) \\
&\leq \frac{1}{2} \sum_{\xi\in\mathbb{N}} \sigma_\xi \left(||v_\xi||^2 + ||u_\xi||^2 \right) \\
&= \sum_{\xi\in\mathbb{N}} \sigma_\xi \\
&= \mathcal{R}_*^\infty(T)
\end{aligned}$$

holds which due to the arbitrariness of $(\bar{e}_\phi)_{\phi\in\mathbb{N}}$ and $(\bar{f}_\phi)_{\phi\in\mathbb{N}}$ can be generalized to

$$\sup_{\substack{(e_\phi)_\phi, (f_\phi)_\phi \\ \text{ONS}}} \sum_{\phi\in\mathbb{N}} |\langle Tf_\phi, e_\phi\rangle| \leq \mathcal{R}_*^\infty(T).$$

□

With this we have everything at hand to finally demonstrate the validity of Theorem 6.12. In doing so, we abridge the parts that are borrowed from the mixed norm case in the previous subsection.

Proof of Theorem 6.12.

(1) *Lim inf inequality:*
Let $(T_\mu)_{\mu\in\mathbb{N}} \subset \mathcal{I}^{2,2}(\Omega,\Sigma)$ be a convergent sequence with respect to the weak operator topology. Then due to the equivalence of convergence concepts on integral operators and their kernels (cf. Lemma 6.6) its limit T_* lies in $\mathcal{I}^{2,2}(\Omega,\Sigma)$ and the inequality holds immediately for $\liminf_{\mu\to\infty} \mathcal{R}_*^{m_\mu,n_\mu}(T_\mu) = \infty$.
Thus, let $\liminf_{\mu\to\infty} \mathcal{R}_*^{m_\mu,n_\mu}(T_\mu) < \infty$. Then there exists a subsequence $(T_{\mu_\phi})_{\phi\in\mathbb{N}} \subseteq (T_\mu)_{\mu\in\mathbb{N}}$ such that

$$\infty > \liminf_{\mu\to\infty} \mathcal{R}_*^{m_\mu,n_\mu}(T_\mu) = \lim_{\phi\to\infty} \mathcal{R}_*^{m_{\mu_\phi},n_{\mu_\phi}}(T_{\mu_\phi}),$$

which implies

$$\sup_{\phi\in\mathbb{N}} \mathcal{R}_*^{m_{\mu_\phi},n_{\mu_\phi}}(T_{\mu_\phi}) < \infty.$$

This, together with the reflexivity of $L^2(\Sigma\times\Omega)$, enables us to apply Banach-Alaoglu and infer the existence of a weakly convergent subsubsequence $(T_{\mu_{\phi_\psi}})_{\psi\in\mathbb{N}}$ with limit T_*. Making use of the in Lemma 6.16 shown lower semicontinuity of $\mathcal{R}_*^\infty$ and involving Remark 6.14 then leads to

$$\begin{aligned}
\liminf_{\mu\to\infty} \mathcal{R}_*^{m_\mu,n_\mu}(T_\mu) &= \lim_{\phi\to\infty} \mathcal{R}_*^{m_{\mu_\phi},n_{\mu_\phi}}(T_{\mu_\phi}) \\
&= \liminf_{\psi\to\infty} \mathcal{R}_*^{m_{\mu_{\phi_\psi}},n_{\mu_{\phi_\psi}}}(T_{\mu_{\phi_\psi}}) \\
&= \liminf_{\psi\to\infty} \mathcal{R}_*^{\infty}(T_{\mu_{\phi_\psi}}) \\
&\geq \mathcal{R}_*^\infty(T_*).
\end{aligned}$$

(2) *Existence of a recovery sequence:*
Let $\widehat{\mathcal{I}}^{2,2}(\Omega,\Sigma)$ be the set of all integral operators in $\mathcal{I}^{2,2}(\Omega,\Sigma)$ with finite number of nonzero singular values, i.e.

$$\widehat{\mathcal{I}}^{2,2}(\Omega,\Sigma) := \left\{T\in\mathcal{I}^{2,2}(\Omega,\Sigma) \,|\, \exists\, \Xi\in\mathbb{N} : \sigma_\xi(T) = 0 \;\forall \xi > \Xi\right\}.$$

Furthermore, suppose that $T\in\mathcal{I}^{2,2}(\Omega,\Sigma)$ is equipped with the singular system $(\sigma_\xi,u_\xi,v_\xi)_\xi$ for $\xi\in\mathbb{N}$. Then, according to Lemma 2.17 the corresponding kernel $t\in L^2(\Sigma\times\Omega)$ can be represented by

$$t(s,r) = \sum_{\xi\in\mathbb{N}} \sigma_\xi u_\xi(s) v_\xi(r)$$

and its truncation $t^{\Xi} \in L^2(\Sigma \times \Omega)$ with

$$t^{\Xi}(s,r) = \sum_{\xi=1}^{\Xi} \sigma_\xi u_\xi(s) v_\xi(r), \quad \Xi \in \mathbb{N}$$

uniquely defines an operator $T^{\Xi} \in \widehat{\mathcal{I}}^{2,2}(\Omega, \Sigma)$. Respecting the positivity of all singular values we realize that this truncated version of T fulfills

$$\begin{aligned}
||T - T^{\Xi}||^2_{\mathcal{I}^{2,2}(\Omega,\Sigma)} &= \left|\left| \sum_{\xi=\Xi+1}^{\infty} \sigma_\xi u_\xi v_\xi \right|\right|^2_{L^2(\Sigma\times\Omega)} \\
&= \sum_{\xi=\Xi+1}^{\infty} \sum_{\xi'=\Xi+1}^{\infty} \sigma_\xi \sigma_{\xi'} \langle u_\xi, u_{\xi'} \rangle_{L^2(\Sigma)} \langle v_\xi, v_{\xi'} \rangle_{L^2(\Omega)} \\
&= \sum_{\xi=\Xi+1}^{\infty} \sigma_\xi^2 \\
&\le \left(\sum_{\xi=\Xi+1}^{\infty} \sigma_\xi \right)^2 .
\end{aligned}$$

Due to the convergence of $(\sigma_\xi)_{\xi\in\mathbb{N}}$ toward zero, this implies that given any $\varepsilon > 0$ we are able to find $\Xi_0 \in \mathbb{N}$ big enough such that

$$||T - T^{\Xi_0}||_{\mathcal{I}^{2,2}(\Omega,\Sigma)} \le \varepsilon.$$

Hence, $\widehat{\mathcal{I}}^{2,2}(\Omega, \Sigma)$ lies dense in $\mathcal{I}^{2,2}(\Omega, \Sigma)$ and according to Remark 2.21 it is sufficient to find a recovery sequence for all $T \in \widehat{\mathcal{I}}^{2,2}(\Omega, \Sigma)$.
So let $T \in \widehat{\mathcal{I}}^{2,2}(\Omega, \Sigma)$ and $(\sigma_\xi, u_\xi, v_\xi)_{\xi\in\{1,\dots,\Xi\}}$ be its associated singular system. Moreover, let $t \in L^2(\Sigma \times \Omega)$ with

$$t(s,r) = \sum_{\xi=1}^{\Xi} \sigma_\xi u_\xi(s) v_\xi(r)$$

denote its corresponding integral kernel. Then, we define the sequence of operators $(T_\mu)_{\mu\in\mathbb{N}} \subset \widehat{\mathcal{I}}^{2,2}(\Omega, \Sigma)$ via the sequence of its inducing kernels $(t_\mu)_{\mu\in\mathbb{N}} \subset L^2(\Sigma \times \Omega)$ which are characterized by

$$t_\mu(s,r) := \sum_{\xi=1}^{\Xi} \sigma_\xi (P_{m_\mu} u_\xi)(s)(Q_{n_\mu} v_\xi)(r).$$

Here, for any $m, n \in \mathbb{N}$, $P_{m_\mu} : L^2(\Sigma) \to L^2(\Sigma)$ and $Q_{n_\mu} : L^2(\Omega) \to L^2(\Omega)$ indicate projections that we want to define via

$$P_m x := \sum_{i=1}^{m} \frac{\int_{\Sigma_i^m} x(s)\,\mathrm{d}s}{|\Sigma_i^m|} \chi_{\Sigma_i^m}, \quad Q_n y := \sum_{j=1}^{n} \frac{\int_{\Omega_j^n} y(r)\,\mathrm{d}r}{|\Omega_j^n|} \chi_{\Omega_j^n}. \tag{6.21}$$

Note that, since Jensen's inequality ensures that for all $x \in L^2(\Sigma)$, $y \in L^2(\Omega)$ with $||x||_{L^2(\Sigma)} \leq 1$, $||y||_{L^2(\Omega)} \leq 1$

$$||P_m x||^2_{L^2(\Sigma)} = \sum_{i=1}^{m} \left| \frac{\int_{\Sigma_i^m} x(s)\,\mathrm{d}s}{|\Sigma_i^m|} \right|^2 |\Sigma_i^m| \leq \sum_{i=1}^{m} \frac{\int_{\Sigma_i^m} |x(s)|^2\,\mathrm{d}s}{|\Sigma_i^m|} |\Sigma_i^m| = ||x||^2_{L^2(\Sigma)} \leq 1, \tag{6.22}$$

$$||Q_n y||^2_{L^2(\Omega)} \leq \sum_{j=1}^{n} \frac{\int_{\Omega_j^n} |y(r)|^2\,\mathrm{d}r}{|\Omega_j^n|} |\Omega_j^n| \leq 1 \tag{6.23}$$

holds, these definitions essentially imply the boundedness of $||P_m||$ and $||Q_n||$ by 1. In fact this boundedness is even independent of the parameters m and n.
With the objective of validating the convergence of $(T_\mu)_{\mu\in\mathbb{N}}$ toward T we now want to scrutinize the behavior of $||T - T_\mu||_{\mathcal{I}^{2,2}(\Omega,\Sigma)}$. To this end we initially ascertain that

$$\begin{aligned} ||T - T_\mu||_{\mathcal{I}^{2,2}(\Omega,\Sigma)} &= \left|\left| \sum_{\xi=1}^{\Xi} \sigma_\xi u_\xi v_\xi - \sum_{\xi=1}^{\Xi} \sigma_\xi P_{m_\mu} u_\xi Q_{n_\mu} v_\xi \right|\right|_{L^2(\Sigma\times\Omega)} \\ &= \left|\left| \sum_{\xi=1}^{\Xi} \sigma_\xi \left(u_\xi (v_\xi - Q_{n_\mu} v_\xi) + (u_\xi - P_{m_\mu} u_\xi) Q_{n_\mu} v_\xi \right) \right|\right|_{L^2(\Sigma\times\Omega)} \\ &\leq \sum_{\xi=1}^{\Xi} \sigma_\xi \, ||u_\xi (v_\xi - Q_{n_\mu} v_\xi) + (u_\xi - P_{m_\mu} u_\xi) Q_{n_\mu} v_\xi||_{L^2(\Sigma\times\Omega)} \\ &\leq \sum_{\xi=1}^{\Xi} \sigma_\xi \left(||v_\xi - Q_{n_\mu} v_\xi||_{L^2(\Omega)} + ||u_\xi - P_{m_\mu} u_\xi||_{L^2(\Sigma)} \right). \end{aligned} \tag{6.24}$$

To examine this expression in more detail we first of all take into account that due to the density of $C_c^\infty(\Omega)$ in $L^2(\Omega)$ we can find for every $\xi \in \{1, \dots, \Xi\}$ and any $\varepsilon > 0$ some $v_\xi^\varepsilon \in C_c^\infty(\Omega)$ that meets $||v_\xi - v_\xi^\varepsilon||_{L^2(\Omega)} \leq \frac{\varepsilon}{6\mathcal{R}_*^\infty(T)}$. Thereby, we can deduce that

$$\begin{aligned} ||v_\xi - Q_{n_\mu} v_\xi||_{L^2(\Omega)} &\leq ||v_\xi - v_\xi^\varepsilon||_{L^2(\Omega)} + ||v_\xi^\varepsilon - Q_{n_\mu} v_\xi^\varepsilon||_{L^2(\Omega)} + ||Q_{n_\mu} v_\xi^\varepsilon - Q_{n_\mu} v_\xi||_{L^2(\Omega)} \\ &\leq \frac{\varepsilon}{6\mathcal{R}_*^\infty(T)} + ||v_\xi^\varepsilon - Q_{n_\mu} v_\xi^\varepsilon||_{L^2(\Omega)} + ||Q_{n_\mu}|| \, ||v_\xi^\varepsilon - v_\xi||_{L^2(\Omega)} \\ &\leq \frac{\varepsilon}{3\mathcal{R}_*^\infty(T)} + ||v_\xi^\varepsilon - Q_{n_\mu} v_\xi^\varepsilon||_{L^2(\Omega)} \end{aligned} \tag{6.25}$$

which projects the remaining estimation problem to functions $v_\xi^\varepsilon \in C_c^\infty(\Omega)$. Taking advantage of this continuity and respecting property (##) we perceive similar to (6.13) that there exists $\overline{\mu}_\xi \in \mathbb{N}$ such that for $r \in \Omega_{j'}^{n_\mu}$, $j' \in \{1, \ldots, n_\mu\}$ and all $\mu \geq \overline{\mu}_\xi$

$$|v_\xi^\varepsilon(r) - Q_{n_\mu} v_\xi^\varepsilon(r)| = \frac{1}{|\Omega_{j'}^{n_\mu}|} \left| \int_{\Omega_{j'}^{n_\mu}} v_\xi^\varepsilon(r) - v_\xi^\varepsilon(\tilde{r}) \, \mathrm{d}\tilde{r} \right| \leq \sup_{\tilde{r} \in \Omega_{j'}^{n_\mu}} |v_\xi^\varepsilon(r) - v_\xi^\varepsilon(\tilde{r})| \leq \frac{\varepsilon}{6\sqrt{|\Omega|}\mathcal{R}_*^\infty(T)}$$

applies. This induces that dependent on the contemplated ξ

$$||v_\xi^\epsilon - Q_{n_\mu} v_\xi^\varepsilon||_{L^2(\Omega)} = \left(\int_\Omega |v_\xi^\varepsilon(r) - (Q_{n_\mu} v_\xi^\varepsilon)(r)|^2 \, \mathrm{d}r \right)^{\frac{1}{2}} \leq \frac{\varepsilon}{6\mathcal{R}_*^\infty(T)}$$

holds for μ big enough. Defining

$$\overline{M} := \max_{\xi \in \{1, \ldots, \Xi\}} \overline{\mu}_\xi$$

we can then infer together with (6.25) that even independent of ξ

$$||v_\xi - Q_{n_\mu} v_\xi||_{L^2(\Omega)} \leq \frac{\varepsilon}{2\mathcal{R}_*^\infty(T)} \tag{6.26}$$

holds as long as $\mu \geq \overline{M}$. Beyond that, exploiting the density of $C_c^\infty(\Sigma)$ in $L^2(\Sigma)$ as well as property (#) we can likewise deduce the existence of some $\widetilde{M} \in \mathbb{N}$ which guarantees the validity of

$$||u_\xi - P_{m_\mu} u_\xi||_{L^2(\Sigma)} \leq \frac{\varepsilon}{2\mathcal{R}_*^\infty(T)} \tag{6.27}$$

for all $\mu \geq \widetilde{M}$ and all $\xi \in \{1, \ldots, \Xi\}$. Combining the results in (6.24), (6.26) and (6.27) as well as the statement in Lemma 6.13 we can therefore conclude that

$$||T - T_\mu||_{\mathcal{I}^{2,2}(\Omega,\Sigma)} \leq \frac{\varepsilon}{\mathcal{R}_*^\infty(T)} \sum_{\xi=}^{\Xi} \sigma_\xi = \varepsilon$$

for $\mu \geq M := \max(\overline{M}, \widetilde{M})$ which due to the arbitrariness of ε implies the (norm) convergence of $(T_\mu)_{\mu \in \mathbb{N}}$ toward T.
With this convergence at hand we now want to verify that $(T_\mu)_{\mu \in \mathbb{N}}$ in fact represents a recovery sequence for T. Therefore, let $(\sigma_\eta^\mu, u_\eta^\mu, v_\eta^\mu)_{\eta \in \mathbb{N}}$ denote the respective singular system to any element T_μ. Then, we can argue from Remark 6.14 and the discrete representability

of T_μ through a matrix $A_{m_\mu,n_\mu} \in \mathbb{R}^{m_\mu \times n_\mu}$ that

$$\begin{aligned}
\mathcal{R}_*^{m_\mu,n_\mu}(T_\mu) &= \mathcal{R}_*^{\infty}(T_\mu) \\
&= \sum_{\eta\in\mathbb{N}} \sigma_\eta^\mu \\
&= \sum_{\eta\in\mathbb{N}} \langle T_\mu v_\eta^\mu, u_\eta^\mu\rangle \\
&= \sum_{\eta\in\mathbb{N}} \left\langle \sum_{\xi=1}^{\Xi} \sigma_\xi \langle Q_{n_\mu} v_\xi, v_\eta^\mu\rangle P_{m_\mu} u_\xi, u_\eta^\mu \right\rangle \\
&= \sum_{\eta\in\mathbb{N}} \sum_{\xi=1}^{\Xi} \sigma_\xi \langle Q_{n_\mu} v_\xi, v_\eta^\mu\rangle \langle P_{m_\mu} u_\xi, u_\eta^\mu\rangle \\
&= \sum_{\eta\in\mathbb{N}} \sum_{\xi=1}^{\Xi} \sigma_\xi \langle v_\xi, v_\eta^\mu\rangle \langle u_\xi, u_\eta^\mu\rangle + \underbrace{\sum_{\eta\in\mathbb{N}} \sum_{\xi=1}^{\Xi} \sigma_\xi \langle Q_{n_\mu} v_\xi - v_\xi, v_\eta^\mu\rangle \langle P_{m_\mu} u_\xi, u_\eta^\mu\rangle}_{(*)} \\
&\quad + \underbrace{\sum_{\eta\in\mathbb{N}} \sum_{\xi=1}^{\Xi} \sigma_\xi \langle v_\xi, v_\eta^\mu\rangle \langle P_{m_\mu} u_\xi - u_\xi, u_\eta^\mu\rangle}_{(**)} .
\end{aligned}$$

Taking a closer look on the expressions in $(*)$ and $(**)$ we perceive with the help of the inequalities of Cauchy-Schwarz and Bessel that

$$\begin{aligned}
\sum_{\eta\in\mathbb{N}} \langle Q_{n_\mu} v_\xi - v_\xi, v_\eta^\mu\rangle \langle P_{m_\mu} u_\xi, u_\eta^\mu\rangle &\le \left[\left(\sum_{\eta\in\mathbb{N}} \langle Q_{n_\mu} v_\xi - v_\xi, v_\eta^\mu\rangle^2 \right) \left(\sum_{\eta\in\mathbb{N}} \langle P_{m_\mu} u_\xi, u_\eta^\mu\rangle^2 \right) \right]^{\frac{1}{2}} \\
&\le ||Q_{n_\mu} v_\xi - v_\xi||_{L^2(\Omega)} \, ||P_{m_\mu} u_\xi||_{L^2(\Sigma)}
\end{aligned}$$

and likewise

$$\sum_{\eta\in\mathbb{N}} \langle P_{m_\mu} u_\xi - u_\xi, u_\eta^\mu\rangle \langle v_\xi, v_\eta^\mu\rangle \le ||P_{m_\mu} u_\xi - u_\xi||_{L^2(\Sigma)}.$$

Recalling the results in (6.22), (6.26) and (6.27) these estimations lead to the insight, that for any $\delta > 0$ there exists a variable $M \in \mathbb{N}$ which ensures that

$$\begin{aligned}
\sum_{\eta\in\mathbb{N}} \langle Q_{n_\mu} v_\xi - v_\xi, v_\eta^\mu\rangle \langle P_{m_\mu} u_\xi, u_\eta^\mu\rangle &\le \frac{\delta}{\mathcal{R}_*^{\infty}(T)}, \\
\sum_{\eta\in\mathbb{N}} \langle P_{m_\mu} u_\xi - u_\xi, u_\eta^\mu\rangle \langle v_\xi, v_\eta^\mu\rangle &\le \frac{\delta}{\mathcal{R}_*^{\infty}(T)}
\end{aligned}$$

for all $\mu \ge M$ and all $\xi \in \{1, \dots, \Xi\}$. This in turn, again with Lemma 6.13 and the arbitrariness of δ, entails the vanishing behavior of both terms, $(*)$ and $(**)$, as soon as μ

tends to infinity. Incorporating the result in Lemma 6.17 we hence eventually preserve

$$\begin{aligned}
\lim_{\mu\to\infty} \mathcal{R}_*^{m_\mu,n_\mu}(T_\mu) &= \lim_{\mu\to\infty} \sum_{\eta\in\mathbb{N}} \sum_{\xi=1}^{\Xi} \sigma_\xi \langle v_\xi, v_\eta^\mu\rangle\langle u_\xi, u_\eta^\mu\rangle \\
&= \lim_{\mu\to\infty} \sum_{\eta\in\mathbb{N}} \langle T v_\eta^\mu, u_\eta^\mu\rangle \\
&\leq \lim_{\mu\to\infty} \sup \left\{ \sum_{\phi\in\mathbb{N}} |\langle T f_\phi, e_\phi\rangle| \,\middle|\, (e_\phi)_{\phi\in\mathbb{N}} \subset L^2(\Sigma), (f_\phi)_{\phi\in\mathbb{N}} \subset L^2(\Omega) \text{ ONSs} \right\} \\
&= \mathcal{R}_*^\infty(T).
\end{aligned}$$

□

With this result we now briefly want to turn toward the consideration of both semi-discrete scenarios which, due to their previously mentioned bijective relation to the spaces $\mathcal{I}_n^{2,2}(\Omega,\Sigma) \subset \mathcal{I}^{2,2}(\Omega,\Sigma)$ and $\mathcal{I}_m^{2,2}(\Omega,\Sigma) \subset \mathcal{I}^{2,2}(\Omega,\Sigma)$ (cf. (6.17) and (6.18)), can in some sense be understood as special cases of the fully continuous scenario. To do so we first of all perceive that this bijective relation via the characterizing elements $t \in L^2(\Sigma;\mathbb{R}^n)$, respectively $t \in L^2(\Omega;\mathbb{R}^m)$, is even an isometric isomorphic one as soon as the artificially introduced and therefore freely selectable spaces Ω, respectively Σ, and their corresponding partitions are chosen properly. This means for example that when considering $\mathcal{I}^{2,2}(\mathbb{R}^n,\Sigma)$ an isometric isomorphic relation to $\mathcal{I}_n^{2,2}(\Omega,\Sigma)$ can be achieved by choosing $\Omega = (0,n)$ and defining its partition via

$$\mathcal{O}_n = \{\Omega_j^n = (j-1,j) \,|\, j = 1,\dots n\}.$$

That is because respecting the definition of inner products in Remark 6.4 in this case we can compute for any $T, L \in \mathcal{I}^{2,2}(\mathbb{R}^n,\Sigma)$ and $\widehat{T},\widehat{L} \in \mathcal{I}_n^{2,2}(\Omega,\Sigma) \subset \mathcal{I}^{2,2}(\Omega,\Sigma)$ which share the same characterizing elements $t,\ell \in L^p(\Sigma;\mathbb{R}^n)$ that

$$\begin{aligned}
\langle \widehat{T},\widehat{L}\rangle_{\mathcal{I}^{2,2}(\Omega,\Sigma)} &= \int_\Sigma \int_\Omega \left(\sum_{j=1}^n t_j(s)\chi_{\Omega_j^n}(r)\right)\left(\sum_{j'=1}^n \ell_{j'}(s)\chi_{\Omega_{j'}^n}(r)\right) \mathrm{d}r\,\mathrm{d}s \\
&= \int_\Sigma \sum_{j=1}^n t_j(s)\ell_j(s)|\Omega_j^n|\,\mathrm{d}s \\
&= \sum_{j=1}^n \int_\Sigma t_j(s)\ell_j(s)\,\mathrm{d}s \\
&= \langle T, L\rangle_{\mathcal{I}^{2,2}(\mathbb{R}^n,\Sigma)}.
\end{aligned}$$

The same holds true for the relation between $\mathcal{I}^{2,2}(\Omega,\mathbb{R}^m)$ and $\mathcal{I}_m^{2,2}(\Omega,\Sigma)$ for an appropriate choice of Σ and $\mathcal{S}_m$ that obeys $|\Sigma_i^m| = 1$ for all $i \in \{1,\dots,m\}$. Noticing that all auxilary statements that were used in the previous proof were stated for arbitrary operators $T \in \mathcal{I}^{2,2}(\Omega,\Sigma)$ this observation suggests that the mentioned assertions equally apply to operators in $\mathcal{I}^{2,2}(\mathbb{R}^n, \Sigma)$ and $\mathcal{I}^{2,2}(\Omega, \mathbb{R}^m)$. This in turn would intend that Theorem 6.12 and its proof can almost directly be transferred to these settings.

Indeed, for the sequence of functionals $(\mathcal{R}_*^{m,n})_{m\in\mathbb{N}}$ with elements $\mathcal{R}_*^{m,n} : \mathcal{I}^{2,2}(\mathbb{R}^n, \Sigma) \to \overline{\mathbb{R}}$ which obey the mapping rule in (6.19) we can state the following corollary.

Corollary 6.18. Suppose that $(\mathcal{S}_m)_{m\in\mathbb{N}}$ is a sequence of partitions of Σ with the following property:

$$(\#) \qquad \max_{i\in\{1,\dots,m\}} \operatorname{diam}(\Sigma_i^m) \xrightarrow{m\to\infty} 0.$$

Let furthermore $\omega_{ij} := |\Sigma_i^m|^{\frac{1}{2}}$ for all $i \in \{1, \dots, m\}$, $j \in \{1, \dots, n\}$ and $(f_\phi)_{\phi\in\{1,\dots,n\}}$ be an orthonormal basis of $\mathbb{R}^n$. Then $(\mathcal{R}_*^{m,n})_{m\in\mathbb{N}}$ Γ-converges with respect to the weak operator topology on $\mathcal{I}^{2,2}(\mathbb{R}^n, \Sigma)$ for $m \to \infty$ to $\mathcal{R}_*^{\infty,n} : \mathcal{I}^{2,2}(\mathbb{R}^n, \Sigma) \to \overline{\mathbb{R}}$ with

$$\mathcal{R}_*^{\infty,n}(T) := \sum_{\phi=1}^{n} \left\langle (T^*T)^{\frac{1}{2}} f_\phi \,,\, f_\phi \right\rangle.$$

In order to make the validity of this assertion clear we first of all want to point to the fact that the statements in Lemma 6.13 - Lemma 6.17 can actually be transmitted to the scenario in which $T \in \mathcal{I}^{2,2}(\mathbb{R}^n, \Sigma)$. The following three remarks pave the way for the confirmation of this conjecture.

Remark 6.19. Let Ω and its partition $\mathcal{O}_n$ be chosen such that $\mathcal{I}_n^{2,2}(\Omega, \Sigma)$ is isometrically isomorph to $\mathcal{I}^{2,2}(\mathbb{R}^n, \Sigma)$, i.e. such that $|\Omega_j^n| = 1$ for all $j \in \{1, \dots, n\}$. Then, for any $\widehat{T} \in \mathcal{I}_n^{2,2}(\Omega, \Sigma)$ with characterizing element $t \in L^2(\Sigma; \mathbb{R}^n)$ we can compute similar to Remark 6.14 that for arbitrary $y \in L^2(\Omega)$

$$\begin{aligned}
(\widehat{T}^*\widehat{T}y)(r) &= \int_\Sigma \sum_{j=1}^{n} t_j(s)\chi_{\Omega_j^n}(r)\,(Ty)(s)\,\mathrm{d}s \\
&= \int_\Sigma \sum_{j=1}^{n} t_j(s)\chi_{\Omega_j^n}(r) \int_\Omega \sum_{j'=1}^{n} t_{j'}(s)\chi_{\Omega_{j'}^n}(\tilde{r})y(\tilde{r})\,\mathrm{d}\tilde{r}\,\mathrm{d}s \\
&= \sum_{j=1}^{n} \left(\int_\Sigma t_j(s) \sum_{j'=1}^{n} t_{j'}(s) \int_{\Omega_{j'}^n} y(\tilde{r})\,\mathrm{d}\tilde{r}\,\mathrm{d}s \right) \chi_{\Omega_j^n}(r)
\end{aligned}$$

applies. Thereby inducing that eigenfunctions $v \in L^2(\Omega)$ of $\widehat{T}^*\widehat{T}$ need to be of the form

$$v(r) = \sum_{j=1}^{n} c_j \chi_{\Omega_j^n}(r)$$

for some $c \in \mathbb{R}^n$ this implies that the corresponding eigenproblem obeys

$$
\begin{aligned}
&\widehat{T}^*\widehat{T}v = \lambda v \\
\Leftrightarrow \quad & \int_\Sigma t_j(s) \sum_{j'=1}^{n} t_{j'}(s) \int_{\Omega_{j'}^n} \sum_{\bar{j}=1}^{n} c_{\bar{j}} \chi_{\Omega_{\bar{j}}^n}(\tilde{r}) \,\mathrm{d}\tilde{r} \,\mathrm{d}s = \lambda c_j \quad \forall j \in \{1, \dots, n\} \\
\Leftrightarrow \quad & \int_\Sigma t_j(s) \sum_{j'=1}^{n} t_{j'}(s) c_{j'} \,\mathrm{d}s = \lambda c_j \quad \forall j \in \{1, \dots, n\} \\
\Leftrightarrow \quad & T^*Tc = \lambda c
\end{aligned}
$$

for $\lambda \in \mathbb{R}$. Here, T denotes the associated operator in $\mathcal{I}^{2,2}(\mathbb{R}^n, \Sigma)$ that shares the characterizing element t. Thus, $\widehat{T} \in \mathcal{I}_n^{2,2}(\Omega, \Sigma)$ and $T \in \mathcal{I}^{2,2}(\mathbb{R}^n, \Sigma)$ own the exact same set of singular values. Note that it is also clear that the number of non-zero elements in this set does not exceed n.

Remark 6.20. Let $T \in \mathcal{I}^{2,2}(\mathbb{R}^n, \Sigma)$ with associated singular system $(\sigma_\xi, u_\xi, v_\xi)_{\xi \in \{1,\dots,n\}}$. Analogously to the proof of Lemma 6.13 we can see that

$$
(T^*T)^{\frac{1}{2}} = \sum_{\xi=1}^{n} \sigma_\xi (v_\xi \otimes v_\xi)
$$

such that for any orthonormal basis $(f_\phi)_{\phi \in \{1,\dots,n\}}$ of $\mathbb{R}^n$ Parseval's identity guarantees that

$$
\mathcal{R}_*^{\infty,n}(T) = \sum_{\phi=1}^{n} \sum_{\xi=1}^{n} \sigma_\xi \langle v_\xi, f_\phi \rangle^2 = \sum_{\xi=1}^{n} \sigma_\xi .
$$

Making use of the previous remark and the statement in Lemma 6.13 regarding operators in $\mathcal{I}^{2,2}(\Omega, \Sigma)$ we therefore deduce that the evaluations of $\mathcal{R}_*^{\infty,n}$ at T and $\mathcal{R}_*^{\infty}$ at $\widehat{T} \in \mathcal{I}_n^{2,2}(\Omega, \Sigma)$, which shares the same characterizing element as T, coincide.

Remark 6.21. Let Ω and its partition $\mathcal{O}_n$ be chosen such that $\mathcal{I}_n^{2,2}(\Omega, \Sigma)$ is isometrically isomorph to $\mathcal{I}^{2,2}(\mathbb{R}^n, \Sigma)$. Suppose that $(T_\gamma)_{\gamma \in \mathbb{N}} \subset \mathcal{I}^{2,2}(\mathbb{R}^n, \Sigma)$ is a weakly convergent sequence with limit $T_* \in \mathcal{I}^{2,2}(\mathbb{R}^n, \Sigma)$ and respective associated characterizing elements $(t_\gamma)_{\gamma \in \mathbb{N}}$ and t_* in $L^2(\Sigma; \mathbb{R}^n)$. Then, for all $y \in \mathbb{R}^n$ and all $x \in L^2(\Sigma)$

$$
\lim_{\gamma \to \infty} \int_\Sigma \left(\sum_{j=1}^{n} y_j (t_\gamma)_j(s) \right) x(s) \,\mathrm{d}s = \int_\Sigma \left(\sum_{j=1}^{n} y_j (t_*)_j(s) \right) x(s) \,\mathrm{d}s
$$

applies. Since this equality especially holds true for vectors of the form

$$
y = \left(\int_{\Omega_j^n} \widehat{y}(r) \,\mathrm{d}r \right)_{j=1,\dots,n}
$$

where $\widehat{y}$ denotes an arbitrary function in $L^2(\Omega)$, we can calculate for the corresponding sequence

$(\widehat{T}_\gamma)_{\gamma\in\mathbb{N}}$ of elements in $\mathcal{I}_n^{2,2}(\Omega,\Sigma)$ which share the characterizing elements $(t_\gamma)_{\gamma\in\mathbb{N}}$ that

$$\begin{aligned}\lim_{\gamma\to\infty}\int_\Sigma(\widehat{T}_\gamma\widehat{y})(s)x(s)\,\mathrm{d}s &= \lim_{\gamma\to\infty}\int_\Sigma\left(\sum_{j=1}^n(t_\gamma)_j(s)\int_{\Omega_j^n}\widehat{y}(r)\,\mathrm{d}r\right)x(s)\,\mathrm{d}s\\ &= \int_\Sigma\left(\sum_{j=1}^n(t_*)_j(s)\int_{\Omega_j^n}\widehat{y}(r)\,\mathrm{d}r\right)x(s)\,\mathrm{d}s\\ &= \int_\Sigma\left(\int_\Omega\sum_{j=1}^n(t_*)_j(s)\chi_{\Omega_j^n}(r)y(r)\,\mathrm{d}r\right)x(s)\,\mathrm{d}s\end{aligned}$$

for all $\widehat{y}\in L^2(\Omega)$, $x\in L^2(\Sigma)$. Hence, the weak convergence of $(T_\gamma)_{\gamma\in\mathbb{N}}\subset\mathcal{I}^{2,2}(\mathbb{R}^n,\Sigma)$ toward T_* induces the weak convergence of its corresponding sequence $(\widehat{T}_\gamma)_{\gamma\in\mathbb{N}}\subset\mathcal{I}_n^{2,2}(\Omega,\Sigma)$ toward the natural transform of T_* in $\mathcal{I}_n(\Omega,\Sigma)$.

Combining these three remarks we can extrapolate fairly easy that an operator $T\in\mathcal{I}^{2,2}(\mathbb{R}^n,\Sigma)$ with singular system $(\sigma_\xi,u_\xi,v_\xi)_{\xi\in\{1,\dots,n\}}$ owns the following properties:

- $\mathcal{R}_*^{\infty,n}(T)=\sum_{\xi=1}^n\sigma_\xi$,
- if $T=EA$ for any $A\in\mathbb{R}^{m\times n}$ its singular values coincide with those of the matrix $\omega\cdot A$, where $\omega\in\mathbb{R}^{m\times n}$ is defined by $\omega_{ij}:=|\Sigma_i^m|^{\frac{1}{2}}$,
- if $(T_\gamma)_{\gamma\in\mathbb{N}}\subset\mathcal{I}^{2,2}(\mathbb{R}^n,\Sigma)$ is a sequence of operators with $T_\gamma\rightharpoonup T$ then

$$\liminf_{\gamma\to\infty}\mathcal{R}_*^{\infty,n}(T_\gamma)\geq\mathcal{R}_*^{\infty,n}(T),$$

  i.e. $\mathcal{R}_*^{\infty,n}$ is lower semi-continuous with respect to the weak topology on $\mathcal{I}^{2,2}(\mathbb{R}^n,\Sigma)$,
- $\mathcal{R}_*^{\infty,n}(T)=\sup\sum_{\phi=1}^r|\langle Tf_\phi,e_\phi\rangle|$, where for $r\leq n$ the supremum is taken over all orthonormal systems $(e_\phi)_{\phi\in\{1,\dots,r\}}$, $(f_\phi)_{\phi\in\{1,\dots,r\}}$ in $L^2(\Sigma)$, respectively $\mathbb{R}^n$.

Since these attributes served as the key arguments in the proof of Theorem 6.12 we are convinced that one only has to follow the same line of argument in order to obtain the result in Corollary 6.18. In fact, regarding the existence of a recovery sequence the consideration of $\mathcal{I}^{2,2}(\mathrm{R}^n,\Sigma)$ even facilitates the proof as the finiteness of the set of non-zero singular values can immediately be presumed. Beyond that the construction of this recovery sequence reduces to the piecewise constant approximation of the singular function in $L^2(\Sigma)$ which simplifies the subsequent estimation.

Now devoting ourselves to the remaining second semi-discrete case we can unsurprisingly observe a very similar behavior of the sequence $(\mathcal{R}_*^{m,n})_{n\in\mathbb{N}}$ whose elements $\mathcal{R}_*^{m,n}:\mathcal{I}^{2,2}(\Omega,\mathbb{R}^m)\to\overline{\mathbb{R}}$ follow the specification in (6.19).

Corollary 6.22. Suppose that $(\mathcal{O}_n)_{n\in\mathbb{N}}$ is a sequence of partitions of Ω with the following property:

$$(\#) \qquad \max_{j\in\{1,\dots,n\}} \operatorname{diam}(\Omega_j^n) \xrightarrow{n\to\infty} 0.$$

Let furthermore $\omega_{ij} := |\Omega_j^n|^{\frac{1}{2}}$ for all $i \in \{1,\dots,m\}$, $j \in \{1,\dots,n\}$ and $(e_i)_{i=\{1,\dots,m\}}$ be an orthonormal basis of $\mathbb{R}^m$. Then $(\mathcal{R}_*^{m,n})_{n\in\mathbb{N}}$ Γ-converges with respect to the weak operator topology on $\mathcal{I}^{2,2}(\Omega,\mathbb{R}^m)$ for $n\to\infty$ to $\mathcal{R}_*^{m,\infty} : \mathcal{I}^{2,2}(\Omega,\mathbb{R}^m) \to \overline{\mathbb{R}}$ with

$$\mathcal{R}_*^{m,\infty}(T) := \sum_{i=1}^{m} \left\langle (TT^*)^{\frac{1}{2}} e_i \,,\, e_i \right\rangle .$$

In order to be confident about this statement we only have to realize that the arguments in Lemma 6.13 and Remark 6.20 regarding the eigenvalues of T^*T obviously transfer to the eigenvalues of TT^*. From that point on it is clear that the assertions in Remark 6.19 - Remark 6.21 can be deduced accordingly and a rigorous proof can once again be obtained in analogy to the proof of Theorem 6.12.

With the results in Theorem 6.12, Corollary 6.18 and Corollary 6.22 we could thus extend our asymptotic understanding of functionals of the form (6.7) to those which include a regularizing nuclear norm. Together with the insights in Proposition 6.5 and 6.7 they confirm with respect to all continuity approaches the Γ-convergence of $\mathcal{E}_\alpha^{m,n}$ toward a limit which preserves the minimizing structure of its corresponding sequence.

6.2.3 Generalized nuclear norm

In the previous two sections we highlighted the asymptotic behavior of the mixed norm and the nuclear norm as soon as space and/or time resolutions of the considered argument matrix become infinitely fine. However, the probably most interesting analysis deals with the behavior of the newly introduced generalized nuclear norm.
Similar to the discrete setting, for this purpose we want to contemplate integral operators that do not operate on $L^2(\Omega)$ and $L^2(\Sigma)$ equipped with their natural Euclidean inner products and norms but resort to alternative versions. Therefore, we first of all formally introduce this new type of function space.

Definition 6.23. Let $\Pi \subset \mathbb{R}^d$ be open. Suppose that $F : L^2(\Pi) \to L^2(\Pi)$ is a bijective, bounded operator that fulfills

$$\text{(i)} \quad \exists c > 0 : \int_\Pi x(w)(Fx)(w)\,\mathrm{d}w \geq c||x||^2_{L^2(\Pi)} \quad \forall x \in L^2(\Pi),$$

$$\text{(ii)} \quad \int_\Pi x(w)(Fy)(w)\,\mathrm{d}w = \int_\Pi (Fx)(w)y(w)\,\mathrm{d}w \quad \forall x,y \in L^2(\Pi).$$

Then, we define $\langle\cdot,\cdot\rangle_F : L^2(\Pi) \times L^2(\Pi) \to \mathbb{R}$ via

$$\langle x, y\rangle_F := \int_\Pi x(w)(Fy)(w)\,\mathrm{d}w$$

and denote by $(L^2(\Pi), \langle\,.\,,\,.\,\rangle_F)$ the Hilbert space which equips $L^2(\Pi)$ with this inner product. Furthermore we indicate the corresponding norm on $(L^2(\Pi), \langle\,.\,,\,.\,\rangle_F)$ by $||\,.\,||_F$, i.e. for all $x \in L^2(\Pi)$

$$||x||_F := \langle x, x\rangle_F^{\frac{1}{2}}.$$

Note that this definition directly implies the existence of some $\tilde{c} > 0$ which guarantees that

$$\tilde{c}||x|| \leq ||x||_F$$

holds for all $x \in L^2(\Pi)$. But in fact, we can even make a stronger statement. Due to the demanded boundedness of F we easily calculate with the help of Hölder's inequality that furthermore there exists some $\hat{c} > 0$ such that for all $x \in L^2(\Pi)$

$$||x||_F = \left(\int_\Pi x(w)(Fx)(w)\,\mathrm{d}w\right)^{\frac{1}{2}} \leq ||x||^{\frac{1}{2}}||Fx||^{\frac{1}{2}} \leq \hat{c}||x||$$

applies. Hence, the norms on $(L^2(\Pi), \langle\,.\,,\,.\,\rangle)$ and $(L^2(\Pi), \langle\,.\,,\,.\,\rangle_F)$ are equivalent. Another property which can directly be deduced from Definition 6.23 is the boundedness of the inverse functional of F: Once again consulting Hölder's inequality we infer from the constraint in (i) that

$$c||F^{-1}x||^2 \leq \int_\Pi (F^{-1}x)(w)x(w)\,\mathrm{d}w \leq ||F^{-1}x||||x||$$

holds true for all $x \in L^2(\Pi)$ and thus confirm that $||F^{-1}|| \leq \frac{1}{c}$.

Just as in the discrete setting the consideration of these modified inner products changes our understanding of the singular value decomposition. Therefore, based on Proposition 4.5, we want to reformulate and specify Definition 2.15.

Definition 6.24. Suppose that $F : L^2(\Pi) \to L^2(\Pi)$ and $G : L^2(P) \to L^2(P)$ are operators which fulfill the requirements in Definition 6.23. Let $T \in L(L^2(P), L^2(\Pi))$ be compact. A sequence $(\sigma_\xi^{F,G}, u_\xi^{F,G}, v_\xi^{F,G})_{\xi\in\mathbb{N}}$ is called F *and* G *induced singular system* of T if

(i) $\sigma_\xi^{F,G} > 0$ for all $\xi \in \mathbb{N}$,

(ii) $((\sigma_\xi^{F,G})^2, u_\xi^{F,G})_{\xi\in\mathbb{N}}$ is an eigensystem corresponding to $TT^{*_{F,G}}$, where $T^{*_{F,G}}$ defines the with respect to $(L^2(\Pi), \langle\,.\,,\,.\,\rangle_F)$ and $(L^2(P), \langle\,.\,,\,.\,\rangle_G)$ adjoint operator to T,

(iii) $v_\xi^{F,G} = \frac{T^{*F,G}\ u_\xi^{F,G}}{||T^{*F,G}\ u_\xi^{F,G}||_G}$ for all $\xi \in \mathbb{N}$.

Note that here the main difference to the classical understanding of a singular system is, that due to the assumption of modified Hilbert spaces the adjoint operator associated with T gets influenced by F and G. Consequently, the operator $TT^{*_{F,G}}$ differs from the Euclidean counterpart and a different singular system is obtained. However, the overall structure of this system is retained.

Remark 6.25. Let $(\sigma_\xi^{F,G}, u_\xi^{F,G}, v_\xi^{F,G})_{\xi\in\mathbb{N}}$ be the F and G induced singular system corresponding to a compact operator $T \in L(L^2(P), L^2(\Pi))$. Then, equivalently to Proposition 2.16 we can deduce, that

$$Tv_\xi^{F,G} = \sigma_\xi^{F,G} u_\xi^{F,G}, \qquad T^{*_{F,G}} u_\xi^{F,G} = \sigma_\xi^{F,G} v_\xi^{F,G}$$

hold and the systems $(u_\xi^{F,G})_{\xi\in\mathbb{N}}$ and $(v_\xi^{F,G})_{\xi\in\mathbb{N}}$ are orthonormal with respect to the inner product induced by F, respectively G. Furthermore, $((\sigma_\xi^{F,G})^2, v_\xi^{F,G})_{\xi\in\mathbb{N}}$ is an eigensystem to $T^{*_{F,G}}T$.

Hence, when considering the F and G induced singular system, we can almost proceed with the usual properties. The only adaption that has to be made concerns all upcoming inner products and all expressions that revert to them. As a consequence we unsurprisingly observe that also for the representation which was introduced in Lemma 2.17 a more general version which adapts to our new understanding of a singular system can be established.

Lemma 6.26. Let $T \in \mathcal{I}^{2,2}(P, \Pi)$ and $(\sigma_\xi^{F,G}, u_\xi^{F,G}, v_\xi^{F,G})_{\xi\in\mathbb{N}}$ be a by F and G induced associated singular system. Then,

$$T = \sum_{\xi\in\mathbb{N}} \sigma_\xi^{F,G} u_\xi^{F,G} \otimes G v_\xi^{F,G}$$

holds.

Proof. Let $(\sigma_\xi^{F,G}, u_\xi^{F,G}, v_\xi^{F,G})_{\xi\in\mathbb{N}}$ denote the singular system induced by F and G and associated with $T \in \mathcal{I}^{2,2}(P, \Pi)$. Then, according to Definition 6.24, $((\sigma_\xi^{F,G})^2, u_\xi^{F,G})_{\xi\in\mathbb{N}}$ is an eigensystem to $TT^{*_{F,G}}$ and for all $\xi \in \mathbb{N}$

$$v_\xi^{F,G} = \frac{T^{*_{F,G}} u_\xi^{F,G}}{||T^{*_{F,G}} u_\xi^{F,G}||_G}.$$

applies. Now realizing that

$$\langle y, G^{-1}T^*Fx\rangle_G = \langle y, T^*Fx\rangle = \langle Ty, Fx\rangle = \langle Ty, x\rangle_F,$$

where T^* designates the Euclidean adjoint operator to T, and defining the system $(e_\xi)_{\xi\in\mathbb{N}}$ via

$$e_\xi := F^{\frac{1}{2}} u_\xi^{F,G},$$

we perceive that

$$TG^{-1}T^*F^{\frac{1}{2}}e_\xi = TT^{*_{F,G}}u_\xi^{F,G} = \left(\sigma_\xi^{F,G}\right)^2 u_\xi^{F,G} = \left(\sigma_\xi^{F,G}\right)^2 F^{-\frac{1}{2}}e_\xi.$$

Consequently,

$$\left(F^{\frac{1}{2}}TG^{-\frac{1}{2}}\right)^* \left(F^{\frac{1}{2}}TG^{-\frac{1}{2}}\right) e_\xi = \left(\sigma_\xi^{F,G}\right)^2 e_\xi$$

holds true and $((\sigma_\xi^{F,G})^2, e_\xi)_{\xi\in\mathbb{N}}$ represents an eigensystem to $(F^{\frac{1}{2}}TG^{-\frac{1}{2}})^*(F^{\frac{1}{2}}TG^{-\frac{1}{2}})$. Since, beyond that we can compute that

$$f_\xi := G^{\frac{1}{2}}v_\xi^{F,G} = \frac{G^{\frac{1}{2}}\left(G^{-1}T^*F\right)F^{-\frac{1}{2}}e_\xi}{\left\langle (G^{-1}T^*F)\,F^{-\frac{1}{2}}e_\xi\,,\,G\,(G^{-1}T^*F)\,F^{-\frac{1}{2}}e_\xi\right\rangle} = \frac{\left(F^{\frac{1}{2}}TG^{-\frac{1}{2}}\right)^* e_\xi}{\left\|\left(F^{\frac{1}{2}}TG^{-\frac{1}{2}}\right)^* e_\xi\right\|}$$

applies, Definition 2.15 suggests that $(\sigma_\xi^{F,G}, e_\xi, f_\xi)_{\xi\in\mathbb{N}}$ represents an Euclidean singular system to $F^{\frac{1}{2}}TG^{-\frac{1}{2}}$. Involving the statement in Lemma 2.17 this implies that

$$F^{\frac{1}{2}}TG^{-\frac{1}{2}} = \sum_{\xi\in\mathbb{N}} \sigma_\xi^{F,G} e_\xi \otimes f_\xi = \sum_{\xi\in\mathbb{N}} \sigma_\xi^{F,G}\left(F^{\frac{1}{2}}u_\xi^{F,G}\right)\otimes\left(G^{\frac{1}{2}}v_\xi^{F,G}\right)$$

and we eventually observe that

$$Ty = TG^{-\frac{1}{2}}(G^{\frac{1}{2}}y) = \sum_{\xi\in\mathbb{N}} \sigma_\xi^{F,G}\left\langle G^{\frac{1}{2}}y, G^{\frac{1}{2}}v_\xi^{F,G}\right\rangle u_\xi^{F,G} = \sum_{\xi\in\mathbb{N}} \sigma_\xi^{F,G}\left\langle y, Gv_\xi^{F,G}\right\rangle u_\xi^{F,G}$$

for all $y \in L^2(P)$. □

With Definition 6.24 and its resulting properties at hand we now want to return to the consideration of the open sets $\Sigma \subset \mathbb{R}^d$ and $\Omega \subset \mathbb{R}^{d'}$ and their partitions $\mathcal{S}_m$ and $\mathcal{O}_n$. Based on them, in order to enable the transition from discrete spaces with modified inner products to continuous ones like in Definition 6.23, we first of all want to introduce operators which allow us to extend the effect of a matrix to general functions in $L^2(\Sigma)$, respectively $L^2(\Omega)$.

Definition 6.27. Let $A \in \mathbb{R}^{m\times m}$ and $B \in \mathbb{R}^{n\times n}$. Suppose that $\mathcal{S}_m = \{\Sigma_1^m, \ldots, \Sigma_m^m\}$ and $\mathcal{O}_n = \{\Omega_1^n, \ldots, \Omega_n^n\}$ are given partitions of Σ and Ω. Then, we define the operators $G_A : L^2(\Sigma) \to L^2(\Sigma)$ and $H_B : L^2(\Omega) \to L^2(\Omega)$ via

$$G_A x := \sum_{i=1}^{m}\sum_{i'=1}^{m} (|\Sigma_i^m||\Sigma_{i'}^m|)^{-\frac{1}{2}} A_{i'i} \int_{\Sigma_i^m} x(s)\,\mathrm{d}s\,\chi_{\Sigma_{i'}^m},$$

$$H_B y := \sum_{j=1}^{n}\sum_{j'=1}^{n} (|\Omega_j^n||\Omega_{j'}^n|)^{-\frac{1}{2}} B_{j'j} \int_{\Omega_j^n} y(r)\,\mathrm{d}r\,\chi_{\Omega_{j'}^n}.$$

Since the operators we have just defined have some special properties that we will make use of in the following, we want to start by putting them on record.

Remark 6.28.

(i) Let $A, \hat{A} \in \mathbb{R}^{m\times m}$. Then we compute, that for all $x \in L^2(\Sigma)$

$$\begin{aligned}\left(G_A \circ G_{\hat{A}}\right) x &= \sum_{i=1}^{m}\sum_{i'=1}^{m}(|\Sigma_i^m||\Sigma_{i'}^m|)^{-\frac{1}{2}}\, A_{i'i}\left(\sum_{i^\circ=1}^{m}|\Sigma_{i^\circ}^m|^{-\frac{1}{2}}|\Sigma_i^m|^{\frac{1}{2}}\hat{A}_{ii^\circ}\int_{\Sigma_{i^\circ}^m} x(s)\,\mathrm{d}s\right)\chi_{\Sigma_{i'}^m}\\ &= \sum_{i^\circ=1}^{m}\sum_{i'=1}^{m}(|\Sigma_{i^\circ}^m||\Sigma_{i'}^m|)^{-\frac{1}{2}}\left(\sum_{i=1}^{m}A_{i'i}\hat{A}_{ii^\circ}\right)\int_{\Sigma_{i^\circ}^m} x(s)\,\mathrm{d}s\,\chi_{\Sigma_{i'}^m}\\ &= G_{A\hat{A}}x,\end{aligned}$$

applies. For $B, \hat{B} \in \mathbb{R}^{n\times n}$ we analogously obtain that

$$\left(H_B \circ H_{\hat{B}}\right) y = H_{B\hat{B}}y$$

for all $y \in L^2(\Omega)$.

(ii) If $A \in \mathbb{R}^{m\times m}$ and $B \in \mathbb{R}^{n\times n}$ are symmetric this property transfers to the operators G_A and H_B, i.e. for all $x, \bar{x} \in L^2(\Sigma)$ and all $y, \bar{y} \in L^2(\Omega)$

$$\langle x, G_A\bar{x}\rangle = \langle G_Ax, \bar{x}\rangle, \quad \langle y, H_B\bar{y}\rangle = \langle H_By, \bar{y}\rangle$$

holds true.

(iii) Let $A \in \mathbb{R}^{m\times m}$, $B \in \mathbb{R}^{n\times n}$. Then, for $x \in L^2(\Sigma)$ and $y \in L^2(\Omega)$ with

$$\begin{aligned}\langle x, \chi_{\Sigma_i^m}\rangle &= \int_{\Sigma_i^m} x(s)\,\mathrm{d}s = 0 \quad \forall i = 1,\dots m,\\ \langle y, \chi_{\Omega_j^n}\rangle &= \int_{\Omega_j^n} y(r)\,\mathrm{d}r = 0 \quad \forall j = 1,\dots,n\end{aligned}$$

the evaluations of G_A and H_B vanish. Accordingly, there exist elements $x \neq 0$ and $y \neq 0$ in $L^2(\Sigma)$, respectively $L^2(\Omega)$, which fulfill

$$\langle x, G_Ax\rangle = 0 = \langle y, H_By\rangle$$

and G_A and G_B can at best be positive *semi*definte. Hence, $\langle\,.\,, G_A.\rangle$ and $\langle\,.\,, H_B\,.\rangle$ do not represent proper inner products on $L^2(\Sigma)$, respectively $L^2(\Omega)$. However, restricting ourselves to with respect to $\mathcal{S}_m$ and $\mathcal{O}_n$ piecewise constant functions $x \in L^2(\Sigma)$ and $y \in L^2(\Omega)$ we can deduce that, together with

$$\langle \mathbf{x}, A\mathbf{x}\rangle \geq c||\mathbf{x}||^2, \quad \langle \mathbf{y}, B\mathbf{y}\rangle \geq \hat{c}||\mathbf{y}||^2 \tag{6.28}$$

for $c, \hat{c} > 0$ and all $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{y} \in \mathbb{R}^n$, also

$$\langle x, G_A x\rangle \geq c||x||^2, \quad \langle y, H_B y\rangle \geq \hat{c}||y||^2$$

holds true.

(iv) Let $A \in \mathbb{R}^{m\times m}$, $B \in \mathbb{R}^{n\times n}$. For piecewise constant functions $x \in L^2(\Sigma)$ and $y \in L^2(\Omega)$ characterized through

$$x := \sum_{i=1}^{m} \mathbf{x}_i |\Sigma_i^m|^{-\frac{1}{2}} \chi_{\Sigma_i^m}, \quad y := \sum_{j=1}^{n} \mathbf{y}_j |\Omega_j^n|^{-\frac{1}{2}} \chi_{\Omega_j^n}$$

with $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{y} \in \mathbb{R}^n$ it pertains that

$$G_A x = \sum_{i=1}^{m} (A\mathbf{x})_i |\Sigma_i^m|^{-\frac{1}{2}} \chi_{\Sigma_i^m}, \quad H_B y = \sum_{j=1}^{n} (B\mathbf{y})_j |\Omega_j^n|^{-\frac{1}{2}} \chi_{\Omega_j^n}$$

and therefore especially

$$G_I x = x, \quad H_I y = y.$$

Keeping the Definition in 6.27 and the precedent characteristics in mind, from now on we want to focus on the consideration of sequences of matrices $(C_m)_{m\in\mathbb{N}} \subset \bigcup_{m\in\mathbb{N}} \mathbb{R}^{m\times m}$ and $(D_n)_{n\in\mathbb{N}} \subset \bigcup_{n\in\mathbb{N}} \mathbb{R}^{n\times n}$ whose elements are symmetric and positive definite. Moreover we want to presume that in some way or another they do behave consistently. To formalize this consistency, we will assert that the by Definition 6.27 corresponding sequences $(G_{C_m})_{m\in\mathbb{N}}$ and $(H_{D_n})_{n\in\mathbb{N}}$ converge pointwise against bounded and bijective limit operators $C : L^2(\Sigma) \to L^2(\Sigma)$ and $D : L^2(\Omega) \to L^2(\Omega)$. Asking for these attributes, we then can infer the transfer of the following properties to C and D.

Lemma 6.29. Suppose that $(\mathcal{S}_m)_{m\in\mathbb{N}}$ and $(\mathcal{O}_n)_{n\in\mathbb{N}}$ are sequences of partitions of Σ and Ω with the following properties:

$$(\#) \qquad \max_{i\in\{1,\dots,m_\mu\}} \operatorname{diam}(\Sigma_i^{m_\mu}) \xrightarrow{\mu\to\infty} 0,$$

$$(\#\#) \qquad \max_{j\in\{1,\dots,n_\mu\}} \operatorname{diam}(\Omega_j^{n_\mu}) \xrightarrow{\mu\to\infty} 0.$$

Let $(C_m)_{m\in\mathbb{N}}$ and $(D_n)_{n\in\mathbb{N}}$ be sequences of symmetric and (6.28) fulfilling positive definite matrices for whose corresponding sequences $(G_{C_m})_{m\in\mathbb{N}}$ and $(H_{D_n})_{n\in\mathbb{N}}$ there exist operators $C : L^2(\Sigma) \to L^2(\Sigma)$ and $D : L^2(\Omega) \to L^2(\Omega)$ with

$$\lim_{m\to\infty} ||G_{C_m} x - Cx|| = 0 \quad \forall x \in L^2(\Sigma), \quad \lim_{n\to\infty} ||H_{D_n} y - Dy|| = 0 \quad \forall y \in L^2(\Omega).$$

Then, C and D are bounded, bijective, symmetric and satisfy

$$\langle x, Cx\rangle \geq c||x||^2, \quad \langle y, Dy\rangle \geq \hat{c}||y||^2$$

for $c, \hat{c} > 0$ and all $x \in L^2(\Sigma)$, $y \in L^2(\Omega)$.

Proof. The boundedness of C and D can be directly deduced from the Banach-Steinhaus Theorem (cf., e.g., [Rudin, 1991, Thrm. 2.5]).
Moreover, since the convergence with respect to the strong operator topology also implies the convergence with respect to the weak one the symmetry of C and D follows directly from the symmetry of G_{C_m} and H_{D_n} mentioned in Remark 6.28 (ii).
Regarding the positive definite property of C and D we first of all want to consider the arbitrary function $x \in L^2(\Sigma)$. Reverting to the projection P_m introduced in (6.21) we then can argue from the statement in Remark 6.28 (iii) that

$$\langle P_m x, G_{C_m} P_m x\rangle \geq c\|P_m x\|^2.$$

Now recalling that we have already shown in (6.27) that under assumption (#) $(P_m x)_{m\in\mathbb{N}}$ converges toward x, we infer that

$$\lim_{m\to\infty} \|P_m x\|^2 = \|x\|^2.$$

On the other hand, we ascertain that Cauchy-Schwarz ensures

$$\begin{aligned}
&|\langle P_m x, G_{C_m} P_m x\rangle - \langle x, Cx\rangle| \\
&\quad \leq |\,\langle P_m x - x, G_{C_m} P_m x\rangle + \langle x, (G_{C_m} - C)P_m x\rangle + \langle x, C(P_m x - x)\rangle\,| \\
&\quad \leq \|P_m x - x\|\|G_{C_m}\|\|P_m\|\|x\| + \|x\|\|(G_{C_m} - C)P_m x\| + \|x\|\|C\|\|P_m x - x\|
\end{aligned}$$

and therefore conclude that together with the pointwise convergence of $(G_{C_m})_{m\in\mathbb{N}}$ to C as well as the boundedness of C, $\sup_{m\in\mathbb{N}} \|P_m\|$ (cf. (6.22)) and $\sup_{m\in\mathbb{N}} \|G_{C_m}\|$ (cf. Banach-Steinhaus Theorem)

$$\lim_{m\to\infty} \langle P_m x, G_{C_m} P_m x\rangle = \langle x, Cx\rangle$$

applies. Consequently, we are convinced that

$$\langle x, Cx\rangle \geq c\|x\|^2$$

holds true for all $x \in L^2(\Sigma)$. Together with the assumption in (##) an equivalent argumentation leads to the corresponding statement for D.
With the validation of this latter property it is also evident, that C and D have to be injective. Hence, in order to prove the bijectivity of both operators it remains to deduce their surjectivity. For this purpose, we again start to focus on $(G_{C_m})_{m\in\mathbb{N}}$ and C: Contemplating any fixed $m \in \mathbb{N}$

the comment in Remark 6.28 (iv) and the invertibility of C_m ensure that for any $x \in L^2(\Sigma)$ there exists some $\overline{\mathbf{x}}_m \in \mathbb{R}^m$ such that

$$G_{C_m}\overline{x}_m = P_m x,$$

for

$$\overline{x}_m := \sum_{i=1}^{m} \overline{\mathbf{x}}_m \chi_{\Sigma_i^m}.$$

Now considering the resulting sequence of piecewise constant functions $(\overline{x}_m)_{m\in\mathbb{N}}$ we first of all find, that due to Remark 6.28 (iii) for every $m \in \mathbb{N}$

$$c\|\overline{x}_m\|^2 \leq \langle \overline{x}_m, G_{C_m}\overline{x}_m\rangle \leq \|\overline{x}_m\|\|G_{C_m}\overline{x}_m\|$$

pertains and consequently together with the boundedness of P_m by 1

$$c\|\overline{x}_m\|^2 \leq \|G_{C_m}\overline{x}_m\|^2 = \|P_m x\|^2 \leq \|x\|^2$$

applies. Thus, $(\overline{x}_m)_{m\in\mathbb{N}}$ is bounded and Banach-Alaoglu guarantees the existence of a weakly converging subsequence whose limit we will denote by $\overline{x} \in L^2(\Sigma)$. For this subsequence we then perceive that, without relabeling, for all $z \in L^2(\Sigma)$

$$\begin{aligned} |\langle G_{C_m}\overline{x}_m - C\overline{x}, z\rangle| &\leq |\langle G_{C_m}(\overline{x}_m - \overline{x}), z\rangle| + |\langle G_{C_m}\overline{x} - C\overline{x}, z\rangle| \\ &\leq |\langle \overline{x}_m - \overline{x}, (G_{C_m} - C)z\rangle| + |\langle \overline{x}_m - \overline{x}, Cz\rangle| + |\langle G_{C_m}\overline{x} - C\overline{x}, z\rangle| \\ &\leq \|\overline{x}_m - \overline{x}\|\|(G_{C_m} - C)z\| + |\langle \overline{x}_m - \overline{x}, Cz\rangle| + |\langle G_{C_m}\overline{x} - C\overline{x}, z\rangle| \end{aligned}$$

holds. Due to the pointwise convergence of $(G_{C_m})_{m\in\mathbb{N}}$ toward C and the weak convergence of $(\overline{x}_m)_{m\in\mathbb{N}}$ to $\overline{x}$ this implies the validity of

$$\lim_{m\to\infty}\langle G_{C_m}\overline{x}_m, z\rangle = \langle C\overline{x}, z\rangle$$

for all $z \in L^2(\Sigma)$. Additionally incorporating the convergence of $(P_m x)_{m\in\mathbb{N}}$ to x then ultimately leads to

$$\langle C\overline{x}, z\rangle = \lim_{m\to\infty}\langle G_{C_m}\overline{x}_m, z\rangle = \lim_{m\to\infty}\langle P_m x, z\rangle = \langle x, z\rangle \quad \forall z \in L^2(\Sigma).$$

Hence, we can confirm the existence of some $\overline{x} \in L^2(\Sigma)$ fulfilling

$$C\overline{x} = x,$$

which due to the arbitrariness of x induces the surjectivity of C. Obviously, again an equivalent reasoning yields the corresponding assertion for D. □

With the proof of these properties we are now confident, that $\langle\, . \, , \, . \,\rangle_C$ and $\langle\, . \, , \, . \,\rangle_D$ define proper inner products on $L^2(\Sigma)$, respectively $L^2(\Omega)$. Thus, it is reasonable to follow Definition 6.24 and consider the singular value decomposition of operators $T \in \mathcal{I}^{2,2}(\Omega, \Sigma)$ with respect to them. Beyond that, according to [Riesz and Sz-Nagy, 1955, §104] the proven characteristics in the previous lemma guarantee that symmetric and positive definite operators $C^{\frac{1}{2}}$ and $D^{\frac{1}{2}}$ exist which due to

$$||C^{\frac{1}{2}}x||^2 = \langle C^{\frac{1}{2}}x, C^{\frac{1}{2}}x\rangle = \langle x, Cx\rangle \leq ||C||\,||x||^2 \quad \forall x \in L^2(\Sigma),$$
$$||D^{\frac{1}{2}}y||^2 \leq ||D||\,||y||^2 \quad \forall y \in L^2(\Omega)$$

inherit the boundedness of C and D. Involving that a nontrivial kernel of $C^{\frac{1}{2}}$ or $D^{\frac{1}{2}}$ would directly lead to a nontrivial kernel of C, respectively D, and that a range of $C^{\frac{1}{2}}$ or $D^{\frac{1}{2}}$ which does not complete $L^2(\Sigma)$, respectively $L^2(\Omega)$, would not allow C, respectively D, to do so, we can furthermore also certify the bijectivity of $C^{\frac{1}{2}}$ and $D^{\frac{1}{2}}$. We shall see, that these attributes will benefit us later on. Nevertheless, to realize their full potential, it will be necessary to make further assumptions on $(C_m)_{m\in\mathbb{N}}$ and $(D_n)_{n\in\mathbb{N}}$. Therefore, we introduce the subsequent premise. Therein, in order to have a summary of all presumptions made, we take up the ones which were already mentioned in Lemma 6.29 and complete them with further ones.

Assumption 6.30. Let $(C_m)_{m\in\mathbb{N}} \subset \bigcup_{m\in\mathbb{N}} \mathbb{R}^{m\times m}$ and $(D_n)_{n\in\mathbb{N}} \subset \bigcup_{n\in\mathbb{N}} \mathbb{R}^{n\times n}$ be sequences of matrices. In the further course of this chapter we want to presume that these were already chosen to satisfy the following properties:

- all elements of both sequences are symmetric and fulfill the positive definite characteristic in (6.28),
- the corresponding sequences of operators $(G_{C_m})_{m\in\mathbb{N}}$ and $(H_{D_n})_{n\in\mathbb{N}}$ converge with respect to the strong operator topology toward operators $C : L^2(\Sigma) \to L^2(\Sigma)$ and $D : L^2(\Omega) \to L^2(\Omega)$,
- the corresponding sequences $\left(G_{C_m^{\frac{1}{2}}}\right)_{m\in\mathbb{N}}$ and $\left(H_{D_n^{-\frac{1}{2}}}\right)_{n\in\mathbb{N}}$ converge with respect to the strong operator topology to $C^{\frac{1}{2}}$, respectively $D^{-\frac{1}{2}}$,
- there exist global constants $c, \bar{c} > 0$ which guarantee that for all $m, n \in \mathbb{N}$

  $$\|C_m x\| \leq c\|x\| \quad \forall x \in \mathbb{R}^m \qquad \text{and} \qquad \|D_n y\| \leq \bar{c}\|y\| \quad \forall y \in \mathbb{R}^n$$

  hold,
- in addition to the positive definite characteristic in (6.28) there exist global constants

$\tilde{c}, \hat{c} > 0$ such that for all $m, n \in \mathbb{N}$

$$\|x\| \leq \tilde{c}\|x\|_{C_m} \quad \forall x \in \mathbb{R}^m \qquad \text{and} \qquad \|y\| \leq \hat{c}\|y\|_{D_n} \quad \forall y \in \mathbb{R}^n$$

apply.

Note that especially the latter two presumed characteristics do not restrict the choices of $(C_m)_{m\in\mathbb{N}}$ and $(D_n)_{n\in\mathbb{N}}$ much further. Although the pointwise convergence of $(G_{C_m})_{m\in\mathbb{N}}$ and $(H_{D_n})_{n\in\mathbb{N}}$ toward the bounded operators C and D, whose associated inner products are equivalent to the Euclidean one, did not directly imply these two properties, we can be sure that we were already quite close.

Within this setting we are now interested in contemplating problems which, in the sense of formulation (6.1), involve regularizers of the form

$$\mathcal{R}(A) = ||A||_{\omega;*_{C_m,D_n}}.$$

In order to address their asymptotic analysis, we again first of all want to check if the statements in section 6.1 apply. Therefore, we need to establish the existence of some $\beta > 0$ which, independent from m and n, ensures that

$$\|A\|_{\overline{\omega};2,2} \leq \beta||A||_{\omega;*_{C_m,D_n}}$$

for all $A \in \mathbb{R}^{m\times n}$. To do so we initially recall the estimation in (6.14): If $(\lambda_\eta, \mathbf{e}_\eta, \mathbf{f}_\eta)_{\eta\in\{1,\ldots,\min(m,n)\}}$ denotes an Euclidean singular system associated with the pointwise product of the arbitrary matrices $\omega, A \in \mathbb{R}^{m\times n}$ and if $\overline{\omega} \in \mathbb{R}^{m\times n}$ is defined via $\overline{\omega}_{ij} := \omega_{ij}^2$, then

$$\|A\|_{\overline{\omega};2,2} \leq \sum_{\eta}^{\min(m,n)} \lambda_\eta$$

holds true. Now furthermore assuming that $(\sigma_\xi^{C_m,D_n}, \mathbf{u}_\xi^{C_m,D_n}, \mathbf{v}_\xi^{C_m,D_n})_{\xi\in\{1,\ldots,\min(m,n)\}}$ represents an according to Proposition 4.5 C_m and D_n induced generalized singular system corresponding to $\omega \cdot A$ we can infer that

$$\omega \cdot A = \sum_{\xi=1}^{\min(m,n)} \sigma_\xi^{C_m,D_n} \mathbf{u}_\xi^{C_m,D_n} \otimes \left(D_n \mathbf{v}_\xi^{C_m,D_n}\right)$$

is fulfilled (cf. (4.8)) and conclude

$$
\begin{aligned}
\|A\|_{\overline{\omega};2,2} &\leq \sum_{\eta=1}^{\min(m,n)} |\langle(\omega \cdot A)\mathbf{f}_\eta, \mathbf{e}_\eta\rangle| \\
&\leq \sum_{\eta=1}^{\min(m,n)} \sum_{\xi=1}^{\min(m,n)} \sigma_\xi^{C_m,D_n} \left| \left\langle \mathbf{u}_\xi^{C_m,D_n}, \mathbf{e}_\eta \right\rangle \right| \left| \left\langle D_n \mathbf{v}_\xi^{C_m,D_n}, \mathbf{f}_\eta \right\rangle \right| \\
&\leq \sum_{\xi=1}^{\min(m,n)} \frac{\sigma_\xi^{C_m,D_n}}{2} \sum_{\eta=1}^{\min(m,n)} \left| \left\langle \mathbf{u}_\xi^{C_m,D_n}, \mathbf{e}_\eta \right\rangle \right|^2 + \left| \left\langle D_n \mathbf{v}_\xi^{C_m,D_n}, \mathbf{f}_\eta \right\rangle \right|^2 \\
&\leq \sum_{\xi=1}^{\min(m,n)} \frac{\sigma_\xi^{C_m,D_n}}{2} \left(\left\| \mathbf{u}_\xi^{C_m,D_n} \right\|^2 + \left\| D_n \mathbf{v}_\xi^{C_m,D_n} \right\|^2 \right).
\end{aligned}
$$

Here we made use of the inequalities of Young and Bessel. Together with the characteristics in Assumption 6.30 and the orthonormality of $(\mathbf{u}_\xi^{C_m,D_n})_\xi$ and $(\mathbf{v}_\xi^{C_m,D_n})_\xi$ with respect to the by C_m, respectively D_n, induced inner product we thus deduce, that independent from m and n for all $A \in \mathbb{R}^{m\times n}$

$$
\|A\|_{\overline{\omega};2,2} \leq \frac{\tilde{c} + \bar{c}\hat{c}}{2} \sum_{\xi=1}^{\min(m,n)} \sigma_\xi^{C_m,D_n} = \frac{\tilde{c} + \bar{c}\hat{c}}{2} \|A\|_{\omega;*_{C_m,D_n}}
$$

applies. With this we are now confident, that even regularizing with the generalized nuclear norm the equi-coercivity of $(\mathcal{E}_\alpha^{m_\mu,n_\mu})_{\mu\in\mathbb{N}}$ as well as the continuity of the associated data fidelity term is ensured. Consequently, in order to understand the limit behavior of $(\mathcal{E}_\alpha^{m_\mu,n_\mu})_{\mu\in\mathbb{N}}$ it remains to contemplate the asymptotics of the generalized nuclear norm.

While studying these, this time we want to concentrate exclusively on the scenario in which both parameters, m and n, simultaneously tend to infinity. In the previous subsection we already saw that both other semi-continuous scenarios can be understood as special cases of this fully continuous one and therefore, quite straight-forward arguments also imply the according convergence behavior for them. This is why we want to omit further details. So, with regard to the definition in (6.9), we are keen to grasp how the functional $\mathcal{R}^{m,n}_{*_{C_m,D_n}} : \mathcal{I}^{2,2}(\Omega, \Sigma) \to \overline{\mathbb{R}}$ characterized through

$$
\mathcal{R}^{m,n}_{*_{C_m,D_n}}(T) = \begin{cases} \sum_{\xi=1}^{\min(m,n)} \sigma_\xi^{C_m,D_n}(\omega \cdot A), & \text{if } T = EA \text{ for some } A \in \mathbb{R}^{m\times n} \\ \infty, & \text{else} \end{cases}
$$

with $\omega \in \mathbb{R}^{m\times n}$ behaves for $m, n \to \infty$. For this purpose we once again want to resort to the single variable notation (m_μ, n_μ) and study its Γ-convergence as μ tends to infinity.

Theorem 6.31. Suppose that $(\mathcal{S}_{m_\mu})_{\mu\in\mathbb{N}}$, $(\mathcal{O}_{n_\mu})_{\mu\in\mathbb{N}}$ are sequences of partitions of Σ and Ω with the following properties:

$$(\#) \qquad \max_{i\in\{1,\dots,m_\mu\}} \operatorname{diam}(\Sigma_i^{m_\mu}) \xrightarrow{\mu\to\infty} 0,$$

$$(\#\#) \qquad \max_{j\in\{1,\dots,n_\mu\}} \operatorname{diam}(\Omega_j^{n_\mu}) \xrightarrow{\mu\to\infty} 0.$$

Assume that $(C_{m_\mu})_{\mu\in\mathbb{N}}$ and $(D_{n_\mu})_{\mu\in\mathbb{N}}$ are sequences of matrices with respect to which the attributes in Assumption 6.30 apply and let furthermore $\omega_{ij} := (|\Sigma_i^{m_\mu}||\Omega_j^{n_\mu}|)^{\frac{1}{2}}$ for all $i \in \{1,\dots,m_\mu\}$, $j \in \{1,\dots,n_\mu\}$ and $(f_\phi)_{\phi\in\mathbb{N}}$ be an orthonormal basis in $(L^2(\Omega), \langle\,.\,,\,.\,\rangle_D)$. Then $(\mathcal{R}^{m_\mu,n_\mu}_{*C_{m_\mu},D_{n_\mu}})_{\mu\in\mathbb{N}}$ Γ-converges with respect to the weak operator topology on $\mathcal{I}^{2,2}(\Omega,\Sigma)$ for $\mu\to\infty$ to $\mathcal{R}^\infty_{*C,D} : \mathcal{I}^{2,2}(\Omega,\Sigma) \to \mathbb{R}$ with

$$\mathcal{R}^\infty_{*C,D}(T) := \sum_{\phi\in\mathbb{N}} \left\langle (T^*T)^{\frac{1}{2}} f_\phi\,,\, f_\phi \right\rangle_D .$$

Similar to the standard nuclear norm case the proof of this theorem is in need of some auxiliary statements and remarks. In order to state these we start by introducing the 'pseudo singular value decomposition' of all discretely representable operators in $\mathcal{I}^{2,2}(\Omega,\Sigma)$.

Remark 6.32. Let $(\sigma_\xi^{C_m,D_n}, \mathbf{u}_\xi^{C_m,D_n}, \mathbf{v}_\xi^{C_m,D_n})_{\xi\in\{1,\dots,\min(m,n)\}}$ be a singular system to any matrix $(\omega\cdot A) \in \mathbb{R}^{m\times n}$ with respect to the inner products in $(\mathbb{R}^m, \langle\,.\,,\,.\,\rangle_{C_m})$ and $(\mathbb{R}^n, \langle\,.\,,\,.\,\rangle_{D_n})$, i.e.

$$\omega\cdot A = \sum_{\xi=1}^{\min(m,n)} \sigma_\xi^{C_m,D_n} \mathbf{u}_\xi^{C_m,D_n} \left(D_n \mathbf{v}_\xi^{C_m,D_n}\right)^T$$

(cf. (4.8)). Let furthermore $T \in \mathcal{I}^{2,2}(\Omega,\Sigma)$ be defined as

$$T := \sum_{\xi}^{\min(m,n)} \sigma_\xi^{C_m,D_n} \left(u_\xi^{C_m,D_n} \otimes H_{D_n} v_\xi^{C_m,D_n}\right) \tag{6.29}$$

where $u_\xi^{C_m,D_n} \in L^2(\Sigma)$ and $v_\xi^{C_m,D_n} \in L^2(\Omega)$ are characterized through

$$u_\xi^{C_m,D_n} := \sum_{i=1}^m \left(\mathbf{u}_\xi^{C_m,D_n}\right)_i |\Sigma_i^m|^{-\frac{1}{2}} \chi_{\Sigma_i^m}, \quad v_\xi^{C_m,D_n} := \sum_{j=1}^n \left(\mathbf{v}_\xi^{C_m,D_n}\right)_j |\Omega_j^n|^{-\frac{1}{2}} \chi_{\Omega_j^n}.$$

Now choosing $\omega_{ij} = (|\Sigma_i^m||\Omega_j^n|)^{\frac{1}{2}}$ and applying the property in Remark 6.28 (iv) we realize that its corresponding integral kernel $t \in L^2(\Sigma\times\Omega)$ obeys

$$t(s,r) = \sum_{\xi}^{\min(m,n)} \sigma_\xi^{C_m,D_n} u_\xi^{C_m,D_n}(s) \left(H_{D_n} v_\xi^{C_m,D_n}\right)(r) = \sum_{i=1}^m \sum_{j=1}^n A_{ij} \chi_{\Sigma_i^m}(s) \chi_{\Omega_j^n}(r)$$

which implies the coincidence of T and EA. Consequently, with (6.29) we found out how any discrete respresentable operator $T = EA$ can be described through the C_m and D_n dependent singular value decomposition of its corresponding matrix $(\omega \cdot A)$. Recalling the in Lemma 6.26 derived representation of general operators in $\mathcal{I}^{2,2}(\Omega, \Sigma)$ through their singular system this suggests to understand $(\sigma_\xi, u_\xi^{C_m,D_n}, v_\xi^{C_m,D_n})_{\xi \in \{1,\dots,\min(m,n)\}}$ as some sort of 'pseudo singular value decomposition' of $T = EA$ with respect to the operators G_{C_m} and H_{D_n}. Although these operators do not fulfill the requirements to define a scalar product on $L^2(\Sigma)$, respectively $L^2(\Omega)$, and thus the definition of a proper singular value decomposition is not feasible, they do own these properties on the subspaces of piecewise constant functions of the form

$$x = \sum_{i=1}^{m} \mathbf{x}_i |\Sigma_i^m|^{-\frac{1}{2}} \chi_{\Sigma_i^m} \quad \text{and} \quad y = \sum_{j=1}^{n} \mathbf{y}_j |\Omega_j^n|^{-\frac{1}{2}} \chi_{\Omega_j^n}$$

with $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{y} \in \mathbb{R}^n$. This is why we can compute that for all $\xi, \psi \in \{1, \dots, \min(m,n)\}$

$$\left\langle u_\xi^{C_m,D_n}, G_{C_m} u_\psi^{C_m,D_n} \right\rangle = \sum_{i=1}^{m} \left(\mathbf{u}_\xi^{C_m,D_n}\right)_i \left(C_m \mathbf{u}_\psi^{C_m,D_n}\right)_i = \left\langle \mathbf{u}_\xi^{C_m,D_n}, \mathbf{u}_\psi^{C_m,D_n} \right\rangle_{C_m} = \delta_{\xi\psi} \quad (6.30)$$

$$\left\langle v_\xi^{C_m,D_n}, H_{D_n} v_\psi^{C_m,D_n} \right\rangle = \sum_{j=1}^{n} \left(\mathbf{v}_\xi^{C_m,D_n}\right)_j \left(D_n \mathbf{v}_\psi^{C_m,D_n}\right)_j = \left\langle \mathbf{v}_\xi^{C_m,D_n}, \mathbf{v}_\psi^{C_m,D_n} \right\rangle_{D_n} = \delta_{\xi\psi} \quad (6.31)$$

applies and even verify that

$$\begin{aligned} T v_\psi^{C_m,D_n} &= \sum_\xi \sigma_\xi^{C_m,D_n} u_\xi^{C_m,D_n} \left\langle v_\psi^{C_m,D_n}, H_{D_n} v_\xi^{C_m,D_n} \right\rangle \\ &= \sum_\xi \sigma_\xi^{C_m,D_n} u_\xi^{C_m,D_n} \sum_{j=1}^{n} \sum_{j'=1}^{n} \left(\mathbf{v}_\psi^{C_m,D_n}\right)_j \left(D_n \mathbf{v}_\xi^{C_m,D_n}\right)_{j'} (|\Omega_j^n||\Omega_{j'}^n|)^{-\frac{1}{2}} \langle \chi_{\Omega_j^n}, \chi_{\Omega_{j'}^n} \rangle \\ &= \sum_\xi \sigma_\xi^{C_m,D_n} u_\xi^{C_m,D_n} \left\langle \mathbf{v}_\psi^{C_m,D_n}, \mathbf{v}_\xi^{C_m,D_n} \right\rangle_{D_n} \\ &= \sigma_\psi^{C_m,D_n} u_\psi^{C_m,D_n}. \end{aligned} \quad (6.32)$$

Hence, $(\sigma_\xi, u_\xi^{C_m,D_n}, v_\xi^{C_m,D_n})_{\xi \in \{1,\dots,\min(m,n)\}}$ indeed fulfills all key properties of a singular system corresponding to T and its designation as the 'pseudo singular value decomposition' with respect to G_{C_m} and H_{D_n} is suitable.

We have thus seen that, in order to be able to consider inner products induced by G_{C_m} and H_{D_n}, temporarily restricting ourselves to the piecewise constant functions in $L^2(\Sigma)$ and $L^2(\Omega)$ can be useful. In the further course of this subsection we will also see that it can be helpful to be able to orthonormalize any arbitrary system of piecewise constant functions in $L^2(\Sigma)$ or $L^2(\Omega)$ with respect to these inner products. Therefore, resorting to the Gram-Schmidt process we want to specify how such orthonormal systems can explicitly be constructed.

Definition 6.33. Let $\Phi \leq m$ and $\Psi \leq n$. Suppose that $(x_\phi)_{\phi\in\{1,\ldots,\Phi\}} \subset L^2(\Sigma)$ and $(y_\psi)_{\psi\in\{1,\ldots,\Psi\}} \subset L^2(\Omega)$ are systems whose elements are piecewise constant with respect to the partitions in $\mathcal{S}_m$, respectively $\mathcal{O}_n$. Then, following the Gram-Schmidt process we recursively construct the systems $(\bar{x}_\phi)_{\phi\in\{1,\ldots,\Phi\}}$ and $(\bar{y}_\psi)_{\psi\in\{1,\ldots,\Psi\}}$ via

$$\bar{x}_1 := \frac{x_1}{\langle x_1, G_{C_m} x_1\rangle}, \quad \tilde{x}_\phi := x_\phi - \sum_{\phi'=1}^{\phi-1} \langle \bar{x}_{\phi'}, G_{C_m} x_\phi\rangle \bar{x}_{\phi'}, \quad \bar{x}_\phi := \frac{\tilde{x}_\phi}{\langle \tilde{x}_\phi, G_{C_m}\tilde{x}_\phi\rangle};$$

$$\bar{y}_1 := \frac{y_1}{\langle y_1, H_{D_n} y_1\rangle}, \quad \tilde{y}_\psi := y_\psi - \sum_{\psi'=1}^{\psi-1} \langle \bar{y}_{\psi'}, H_{D_n} y_\psi\rangle \bar{y}_{\psi'}, \quad \bar{y}_\psi := \frac{\tilde{y}_\psi}{\langle \tilde{y}_\psi, H_{D_n}\tilde{y}_\psi\rangle}.$$

Due to the positive definiteness of G_{C_m} and H_{D_n} restricted to the space of piecewise constant functions these are then orthonormal with respect to $\langle\,.\,, G_{C_m}\,.\,\rangle$ and $\langle\,.\,, H_{D_n}\,.\,\rangle$.

In the following our special interest will be devoted to orthonormal systems, which emerge by means of this process from systems of the form $(P_m x_\phi)_{\phi\in\{1,\ldots,\Phi\}}$ and $(Q_n y_\psi)_{\psi\in\{1,\ldots,\Psi\}}$. Here we revert to the previously defined projections P_m and Q_n and contemplate systems $(x_\phi)_{\phi\in\{1,\ldots,\Phi\}}$ and $(y_\psi)_{\psi\in\{1,\ldots,\Psi\}}$ which are orthonormal in $(L^2(\Sigma), \langle\,.\,,\,.\,\rangle_C)$, respectively $(L^2(\Omega), \langle\,.\,,\,.\,\rangle_D)$. Due to the structure of P_m and Q_n as well as the strong convergence of $(C_m)_{m\in\mathbb{N}}$ and $(D_n)_{n\in\mathbb{N}}$ in this particular case a special behavior can be observed.

Lemma 6.34. Let $P_m : L^2(\Sigma) \to L^2(\Sigma)$ and $Q_n : L^2(\Omega) \to L^2(\Omega)$ be defined as in (6.21). Suppose that $(x_\phi)_{\phi\in\{1,\ldots,\Phi\}}$ and $(y_\psi)_{\psi\in\{1,\ldots,\Psi\}}$ are orthonormal systems in $(L^2(\Sigma), \langle\,.\,,\,.\,\rangle_C)$, respectively $(L^2(\Omega), \langle\,.\,,\,.\,\rangle_D)$. Let furthermore $(\bar{x}^m_\phi)_{\psi\in\{1,\ldots,\Phi\}}$ and $(\bar{y}^n_\psi)_{\psi\in\{1,\ldots,\Psi\}}$ denote the from $(P_m x_\phi)_{\psi\in\{1,\ldots,\Phi\}}$ and $(Q_n y_\psi)_{\psi\in\{1,\ldots,\Psi\}}$ via Definition 6.33 constructed orthonormalized systems. Then, for all $\phi \in \{1,\ldots,\Phi\}$ and all $\psi \in \{1,\ldots,\Psi\}$

$$\lim_{m\to\infty} \|\bar{x}^m_\phi - x_\phi\| = 0, \quad \lim_{n\to\infty} \|\bar{y}^n_\psi - y_\psi\| = 0$$

applies.

Since the proof of this lemma follows the same line of argument for both types of systems, the ones in $L^2(\Sigma)$ and the ones in $L^2(\Omega)$, we limit ourselves to exemplifying its validity for systems $(P_m x_\phi)_{\phi\in\{1,\ldots,\Phi\}} \subset L^2(\Sigma)$.

Proof. Let $(\bar{x}^m_\phi)_{\phi\in\{1,\ldots,\Phi\}}$ be the through Definition 6.33 orthonormalized system which emerged from $(P_m x_\phi)_{\phi\in\{1,\ldots,\Phi\}}$, where $(x_\phi)_{\phi\in\{1,\ldots,\Phi\}}$ forms an orthonormal system in $(L^2(\Sigma), \langle\,.\,,\,.\,\rangle_C)$. Furthermore, suppose that $(\tilde{x}^m_\phi)_{\phi\in\{1,\ldots,\Phi\}}$ denotes the in Definition 6.33 mentioned intermediate orthogonal system before normalization. Then, we want to verify the assertion by validating the following two substatements for all $\phi \in \{1,\ldots,\Phi\}$:

$$\text{(i)} \lim_{m\to\infty} \|\tilde{x}^m_\phi - x_\phi\| = 0, \qquad \text{(ii)} \lim_{m\to\infty} \langle \tilde{x}^m_\phi, G_{C_m}\tilde{x}^m_\phi\rangle = 1 = \|x_\phi\|_C.$$

In doing so, we first of all want to turn to the one in (ii). Assuming that for any $\phi \in \{1, \ldots, \Phi\}$ the convergence of $(\tilde{x}_\phi^m)_{m\in\mathbb{N}}$ toward x_ϕ was already shown, we estimate that

$$\begin{aligned} |\langle \tilde{x}_\phi^m, G_{C_m} \tilde{x}_\phi^m \rangle - \|x_\phi\|_C| \leq & |\langle \tilde{x}_\phi^m - x_\phi, G_{C_m} \tilde{x}_\phi^m \rangle| + |\langle x_\phi, (G_{C_m} - C)\, \tilde{x}_\phi^m \rangle| \\ & + |\langle x_\phi, C\left(\tilde{x}_\phi^m - x_\phi\right)\rangle| \\ \leq & \|\tilde{x}_\phi^m - x_\phi\| \|G_{C_m}\| \|\tilde{x}_\phi^m\| + \|x_\phi\| \|(G_{C_m} - C)\, \tilde{x}_\phi^m\| \\ & + \|x_\phi\| \|C\| \|\tilde{x}_\phi^m - x_\phi\|. \end{aligned}$$

Now, involving that according to the Banach-Steinhaus Theorem [Rudin, 1991, Thrm. 2.5] the strong convergence of $(G_{C_m})_{m\in\mathbb{N}}$ implies the boundedness of $\sup_{m\in\mathbb{N}} \|G_{C_m}\|$ we can argue from the convergence of $(\tilde{x}_\phi^m)_{m\in\mathbb{N}}$ as well as the strong convergence of $(G_{C_m})_{m\in\mathbb{N}}$ to C that for any $\delta > 0$

$$|\langle \tilde{x}_\phi^m, G_{C_m} \tilde{x}_\phi^m \rangle - \|x_\phi\|_C| \leq \delta$$

applies as m is big enough. Hence, due to the arbitrariness of δ the convergence of $(\langle \tilde{x}_\phi^m, G_{C_m} \tilde{x}_\phi^m \rangle)_m$ toward 1 is confirmed.

Regarding the statement in (i) we want to apply the concept of induction. Therefore, we initially realize that

$$\|\tilde{x}_1^m - x_1\| = \|P_m x_1 - x_1\|.$$

Recalling that we have already seen in (6.27) that the latter expression can become arbitrary small as soon as m is chosen big enough this attests the norm convergence of $\tilde{x}_1^m$ toward x_1. Now, assuming that for all $\phi \in \{1, \ldots, \phi' - 1\}$ with $2 \leq \phi' \leq \Phi$ $(\tilde{x}_\phi^m)_{m\in\mathbb{N}}$ already converges toward x_ϕ, we can deduce that, as previously shown, also

$$\lim_{m\to\infty} \langle \tilde{x}_\phi^m, G_{C_m} \tilde{x}_\phi^m \rangle = 1$$

applies as $\phi < \phi'$. Consequently, we can guarantee the existence of some $\hat{c} > 0$ which fulfills

$$\frac{\|\tilde{x}_\phi^m\|}{|\langle \tilde{x}_\phi^m, G_{C_m} \tilde{x}_\phi^m \rangle|} \leq \hat{c}$$

for all $\phi \in \{1, \dots, \phi' - 1\}$ and infer that

$$\begin{aligned}
\|\tilde{x}_{\phi'}^m - x_{\phi'}\| &\le \|P_m x_{\phi'} - x_{\phi'}\| + \sum_{\phi=1}^{\phi'-1} |\langle \tilde{x}_\phi^m, G_{C_m} P_m x_{\phi'}\rangle| \frac{\|\tilde{x}_\phi^m\|}{|\langle \tilde{x}_\phi^m, G_{C_m} \tilde{x}_\phi^m\rangle|} \\
&\le \|P_m x_{\phi'} - x_{\phi'}\| + \hat{c} \sum_{\phi=1}^{\phi'-1} \Big(|\langle x_\phi, C x_{\phi'}\rangle| + |\langle x_\phi, (G_{C_m} - C)\, x_{\phi'}\rangle| \\
&\quad + |\langle x_\phi, G_{C_m} (P_m x_{\phi'} - x_{\phi'})\rangle| + |\langle \tilde{x}_\phi^m - x_\phi, G_{C_m} P_m x_{\phi'}\rangle| \Big) \\
&\le \|P_m x_{\phi'} - x_{\phi'}\| + \hat{c} \sum_{\phi=1}^{\phi'-1} \Big(\|x_\phi\| \|(G_{C_m} - C)\, x_{\phi'}\| \\
&\quad + \|x_\phi\| \|G_{C_m}\| \|P_m x_{\phi'} - x_{\phi'}\| + \|\tilde{x}_\phi^m - x_\phi\| \|G_{C_m}\| \|P_m\| \|x_{\phi'}\| \Big).
\end{aligned}$$

Once again referring to the result in (6.27) and additionally incorporating the strong convergence of $(G_{C_m})_{m\in\mathbb{N}}$ toward C as well as the in (6.22) shown boundedness of P_m we are convinced that for any $\varepsilon > 0$ we can find some $M \in \mathbb{N}$ which ensures

$$\|\tilde{x}_{\phi'}^m - x_{\phi'}\| \le \varepsilon$$

for $m \ge M$. Thus, the general convergence of $(\tilde{x}_\phi^m)_{m\in\mathbb{N}}$ to x_ϕ is proven and we deduce

$$\lim_{m\to\infty} \bar{x}_\phi^m = \lim_{m\to\infty} \frac{\tilde{x}_\phi^m}{\langle \tilde{x}_\phi^m, G_{C_m} \tilde{x}_\phi^m\rangle} = x_\phi$$

for all $\phi \in \{1, \dots, \Phi\}$. □

Coming back to more general functions and operators we now want to derive two alternative representations for $\mathcal{R}_{*C,D}^\infty$ which will prove useful later on. This happens, in consideration of the new modified inner products, equivalently to the statements in Lemma 6.13 and Lemma 6.17.

Lemma 6.35. Let $T \in \mathcal{I}^{2,2}(\Omega, \Sigma)$ and $(\sigma_\xi^{C,D})_{\xi\in\mathbb{N}}$ be the sequence of corresponding singular values with respect to the inner products in $(L^2(\Sigma), \langle\,.\,,\,.\,\rangle_C)$ and $(L^2(\Omega), \langle\,.\,,\,.\,\rangle_D)$. Then, it holds that

$$\mathcal{R}_{*C,D}^\infty(T) = \sum_{\xi\in\mathbb{N}} \sigma_\xi^{C,D}.$$

Proof. Let $T \in \mathcal{I}^{2,2}(\Omega, \Sigma)$ and $(\sigma_\xi^{C,D}, u_\xi^{C,D}, v_\xi^{C,D})_{\xi\in\mathbb{N}}$ be an associated singular system with respect to C and D. Then, recalling the representation of T through this singular system (cf.

Lemma 6.26) we compute that for all $x \in L^2(\Sigma)$, $y \in L^2(\Omega)$

$$\langle x, Ty\rangle_C = \sum_{\xi\in\mathbb{N}} \sigma_\xi^{C,D} \left\langle y, v_\xi^{C,D}\right\rangle_D \left\langle x, u_\xi^{C,D}\right\rangle_C = \left\langle \sum_{\xi\in\mathbb{N}} \sigma_\xi^{C,D} \left\langle x, u_\xi^{C,D}\right\rangle_C v_\xi^{C,D}, y \right\rangle_D$$

applies, such that we can deduce that

$$T^*T = \sum_{\xi\in\mathbb{N}} \left(\sigma_\xi^{C,D}\right)^2 \left(v_\xi^{C,D} \otimes D v_\xi^{C,D}\right).$$

Now, following the same line of argument as in Lemma 6.13 we can verify that

$$(T^*T)^{\frac{1}{2}} = \sum_{\xi\in\mathbb{N}} \left(\sigma_\xi^{C,D}\right) \left(v_\xi^{C,D} \otimes D v_\xi^{C,D}\right)$$

and therefore

$$\mathcal{R}^\infty_{*_{C,D}}(T) = \sum_{\xi\in\mathbb{N}} \sigma_\xi^{C,D}.$$

□

Lemma 6.36. For all $T \in \mathcal{I}^{2,2}(\Omega, \Sigma)$ it holds, that

$$\mathcal{R}^\infty_{*_{C,D}}(T) = \sup \sum_{\phi\in\mathbb{N}} |\langle T f_\phi, e_\phi\rangle_C|$$

where the supremum is taken over all orthonormal systems $(e_\phi)_{\phi\in\mathbb{N}}$, $(f_\phi)_{\phi\in\mathbb{N}}$ in $(L^2(\Sigma), \langle\, .\,, .\,\rangle_C)$, respectively $(L^2(\Omega), \langle\, .\,, .\,\rangle_D)$.

Proof. For proving this statement we can once again stick to the reasoning pursued in the standard nuclear norm case. Substituting the Euclidean inner products with the respective ones in $(L^2(\Sigma), \langle\, .\,, .\,\rangle_C)$ and $(L^2(\Omega), \langle\, .\,, .\,\rangle_D)$ and using the representation in Lemma 6.26, following the proof of Lemma 6.17 yields the assertion. □

Note that this latter representation of $\mathcal{R}^\infty_{*_{C,D}}$ can easily be transformed to work on Euclidean orthonormal systems in $L^2(\Sigma)$ and $L^2(\Omega)$.

Remark 6.37. Let $T \in \mathcal{I}^{2,2}(\Omega, \Sigma)$ and $(e_\phi)_{\phi\in\mathbb{N}}$, $(f_\phi)_{\phi\in\mathbb{N}}$ be arbitrary orthonormal systems in $(L^2(\Sigma), \langle\, .\,, .\,\rangle_C)$, respectively $(L^2(\Omega), \langle\, .\,, .\,\rangle_D)$. Then, due to the bijectivity of $C^{\frac{1}{2}}$ and $D^{\frac{1}{2}}$ we can find a bijective relation to systems $(\hat{e}_\phi)_{\phi\in\mathbb{N}} \subset L^2(\Sigma)$ and $(\hat{f}_\phi)_{\phi\in\mathbb{N}} \subset L^2(\Omega)$ whose elements fulfill

$$e_\phi = C^{-\frac{1}{2}} \hat{e}_\phi, \quad f_\phi = D^{-\frac{1}{2}} \hat{f}_\phi.$$

Simultaneously, the symmetry of $C^{\frac{1}{2}}$ and $D^{\frac{1}{2}}$ guarantees that these newly defined systems satisfy

$$\langle \hat{e}_\phi, \hat{e}_\psi \rangle = \langle C^{\frac{1}{2}} e_\phi, C^{\frac{1}{2}} e_\psi \rangle = \langle e_\phi, e_\psi \rangle_C = \delta_{\phi\psi}, \tag{6.33}$$

$$\langle \hat{f}_\phi, \hat{f}_\psi \rangle = \langle D^{\frac{1}{2}} f_\phi, D^{\frac{1}{2}} f_\psi \rangle = \langle f_\phi, f_\psi \rangle_D = \delta_{\phi\psi}. \tag{6.34}$$

and we perceive that for any $T \in \mathcal{I}^{2,2}(\Omega, \Sigma)$ the identity in Lemma 6.36 can be modified to

$$\mathcal{R}^{\infty}_{*C,D}(T) = \sup \sum_{\phi \in \mathbb{N}} \left| \left\langle T D^{-\frac{1}{2}} f_\phi, C^{\frac{1}{2}} e_\phi \right\rangle \right|$$

where now the supremum is taken over all *Euclidean* orthonormal systems in $L^2(\Sigma)$, respectively $L^2(\Omega)$.

In fact a very similar statement can be derived for $\mathcal{R}^{m,n}_{*C_m,D_n}$ restricted to operators which have a matrix representation.

Lemma 6.38. Let $T \in \mathcal{I}^{2,2}(\Omega, \Sigma)$ be an operator which can be represented by a matrix $A \in \mathbb{R}^{m \times n}$, i.e. $T = EA$. Then,

$$\mathcal{R}^{m,n}_{*C_m,D_n}(T) = \sup \sum_{\phi \in \mathbb{N}} \left| \left\langle T H_{D_n^{-\frac{1}{2}}} f_\phi, G_{C_m^{\frac{1}{2}}} e_\phi \right\rangle \right|.$$

Here the supremum was taken over all orthonormal systems $(e_\phi)_{\phi \in \mathbb{N}}$, $(f_\phi)_{\phi \in \mathbb{N}}$ in $L^2(\Sigma)$, respectively $L^2(\Omega)$, both with respect to the Euclidean inner product.

Proof. Let $T \in \mathcal{I}^{2,2}(\Omega, \Sigma)$ be an operator which satisfies $T = EA$ for any matrix $A \in \mathbb{R}^{m \times n}$ and $\omega \in \mathcal{R}^{m \times n}$ be defined through $\omega_{ij} := (|\Sigma_i^m||\Omega_j^n|)^{\frac{1}{2}}$. Furthermore, let $(\sigma_\xi^{C_m,D_n}, \mathbf{u}_\xi^{C_m,D_n}, \mathbf{v}_\xi^{C_m,D_n})_\xi$ with $\xi \in \{1, \ldots, \min(m,n)\}$ denote a singular system corresponding to $\omega \cdot A$ which resorts to the structures in $(\mathbb{R}^m, \langle\, .\, ,\, .\, \rangle_{C_m})$ and $(\mathbb{R}^n, \langle\, .\, ,\, .\, \rangle_{D_n})$. Then, Remark 6.32 guarantees that

$$T = \sum_{\xi=1}^{\min(m,n)} \sigma_\xi^{C_m,D_n} \left(u_\xi^{C_m,D_n} \otimes (H_{D_n}) v_\xi^{C_m,D_n} \right)$$

for

$$u_\xi^{C_m,D_n} := \sum_{i=1}^{m} \left(\mathbf{u}_\xi^{C_m,D_n} \right)_i |\Sigma_i^m|^{-\frac{1}{2}} \chi_{\Sigma_i^m}, \quad v_\xi^{C_m,D_n} := \sum_{j=1}^{n} \left(\mathbf{v}_\xi^{C_m,D_n} \right)_j |\Omega_j^n|^{-\frac{1}{2}} \chi_{\Omega_j^n}.$$

Now defining the systems $(\hat{e}_\xi)_{\xi \in \{1,\ldots,\min(m,n)\}} \subset L^2(\Sigma)$ and $(\hat{f}_\xi)_{\xi \in \{1,\ldots,\min(m,n)\}} \subset L^2(\Omega)$ via

$$\hat{e}_\xi := G_{C_m^{\frac{1}{2}}} u_\xi^{C_m,D_n}, \quad \hat{f}_\xi := H_{D_n^{\frac{1}{2}}} v_\xi^{C_m,D_n}$$

we realize that as in (6.33) and (6.34) the symmetry of $G_{C_m^{\frac{1}{2}}}$ and $H_{D_n^{\frac{1}{2}}}$ and the properties in (6.30) and (6.31) ensure their Euclidean orthonormality. Thus, additionally involving the characteristics in (6.32) and Remark 6.28 (iv) we can compute that

$$
\begin{aligned}
\sup_{\substack{(e_\phi)_\phi,(f_\phi)_\phi \\ \text{eucl. ONS}}} \sum_{\phi\in\mathbb{N}} \left| \left\langle TH_{D_n^{-\frac{1}{2}}} f_\phi, G_{C_m^{\frac{1}{2}}} e_\phi \right\rangle \right| &\geq \sum_{\xi=1}^{\min(m,n)} \left| \left\langle TH_{D_n^{-\frac{1}{2}}} \hat{f}\xi, G_{C_m^{\frac{1}{2}}} \hat{e}\xi \right\rangle \right| \\
&= \sum_{\xi=1}^{\min(m,n)} \left| \left\langle Tv_\xi^{C_m,D_n}, G_{C_m} u_\xi^{C_m,D_n} \right\rangle \right| \\
&= \sum_{\xi=1}^{\min(m,n)} \left| \left\langle \sigma_\xi^{C_m,D_n} u_\xi^{C_m,D_n}, G_{C_m} u_\xi^{C_m,D_n} \right\rangle \right| \\
&= \sum_{\xi=1}^{\min(m,n)} \sigma_\xi^{C_m,D_n} \\
&= \mathcal{R}^{m,n}_{*C_m,D_n}(T).
\end{aligned}
$$

On the other hand, letting $(\tilde{e}_\psi)_{\psi\in\mathbb{N}} \subset L^2(\Sigma)$ and $(\tilde{f}_\psi)_{\psi\in\mathbb{N}} \subset L^2(\Omega)$ be arbitrary Euclidean orthonormal systems we deduce with Young's and Bessel's inequality that

$$
\begin{aligned}
&\sum_{\psi\in\mathbb{N}} \left| \left\langle TH_{D_n^{-\frac{1}{2}}} \tilde{f}_\psi, G_{C_m^{\frac{1}{2}}} \tilde{e}_\psi \right\rangle \right| \\
&\quad= \sum_{\psi\in\mathbb{N}} \left| \left\langle \sum_{\xi=1}^{\min(m,n)} \sigma_\xi^{C_m,D_n} u_\xi^{C_m,D_n} \left\langle H_{D_n^{-\frac{1}{2}}} \tilde{f}_\psi, H_{D_n} v_\xi^{C_m,D_n} \right\rangle, G_{C_m^{\frac{1}{2}}} \tilde{e}_\psi \right\rangle \right| \\
&\quad= \sum_{\psi\in\mathbb{N}} \left| \sum_{\xi=1}^{\min(m,n)} \sigma_\xi^{C_m,D_n} \left\langle \tilde{f}_\psi, H_{D_n^{\frac{1}{2}}} v_\xi^{C_m,D_n} \right\rangle \left\langle u_\xi^{C_m,D_n}, G_{C_m^{\frac{1}{2}}} \tilde{e}_\psi \right\rangle \right| \\
&\quad\leq \frac{1}{2} \sum_{\psi\in\mathbb{N}} \sum_{\xi=1}^{\min(m,n)} \sigma_\xi^{C_m,D_n} \left(\left| \left\langle \tilde{f}_\psi, H_{D_n^{\frac{1}{2}}} v_\xi \right\rangle \right|^2 + \left| \left\langle u_\xi^{C_m,D_n}, G_{C_m^{\frac{1}{2}}} \tilde{e}_\psi \right\rangle \right|^2 \right) \\
&\quad\leq \frac{1}{2} \sum_{\xi=1}^{\min(m,n)} \sigma_\xi^{C_m,D_n} \left(\left\| H_{D_n^{\frac{1}{2}}} v_\xi^{C_m,D_n} \right\|^2 + \left\| G_{C_m^{\frac{1}{2}}} u_\xi^{C_m,D_n} \right\|^2 \right) \\
&\quad= \sum_{\xi=1}^{\min(m,n)} \sigma_\xi^{C_m,D_n} \\
&\quad= \mathcal{R}^{m,n}_{*C_m,D_n}(T)
\end{aligned}
$$

which due to the arbitrariness of $(\tilde{e}_\psi)_{\psi\in\mathbb{N}}$ and $(\tilde{f}_\psi)_{\psi\in\mathbb{N}}$ generalizes to

$$
\sup_{\substack{(e_\phi)_\phi,(f_\phi)_\phi \\ \text{eucl. ONS}}} \sum_{\phi\in\mathbb{N}} \left| \left\langle TH_{D_n^{-\frac{1}{2}}} f_\phi, G_{C_m^{\frac{1}{2}}} e_\phi \right\rangle \right| \leq \mathcal{R}^{m,n}_{*C_m,D_n}(T).
$$

□

With Lemma 6.36 and Lemma 6.38 we thus saw, that exploiting an operators representation through its associated singular system we can find alternative options which characterize its generalized nuclear norm by resorting to maximizing orthonormal systems. Another useful statement which makes use of the same kind of argumentation but on the level of matrices is the following.

Lemma 6.39. Let $\Phi \leq \min(m,n)$ and $(\lambda_\phi)_{\phi\in\{1,\ldots,\Phi\}} \subset \mathbb{R}^+$. Suppose that $(\mathbf{e}_\phi)_{\phi\in\{1,\ldots,\Phi\}}$ and $(\mathbf{f}_\phi)_{\phi\in\{1,\ldots,\Phi\}}$ are orthonormal systems in $(\mathbb{R}^m, \langle\,.\,,\,.\,\rangle_{C_m})$, respectively $(\mathbb{R}^n, \langle\,.\,,\,.\,\rangle_{D_n})$. Then, for a matrix $A \in \mathbb{R}^{m\times n}$ the identity $A = \sum_{\phi=1}^{\Phi} \lambda_\phi \mathbf{e}_\phi \otimes D_n \mathbf{f}_\phi$ implies that

$$\|A\|_{*_{C_m,D_n}} = \sum_{\phi=1}^{\Phi} \lambda_\phi.$$

Proof. Considering the representation of $A \in \mathbb{R}^{m\times n}$ through its by C_m and D_n induced singular system $(\sigma_\xi^{C_m,D_n}, \mathbf{u}_\xi^{C_m,D_n}, \mathbf{v}_\xi^{C_m,D_n})_{\xi\in\{1,\ldots,\min(m,n)\}}$, i.e. respecting that

$$A = \sum_{\xi=1}^{\min(m,n)} \sigma_\xi^{C_m,D_n} \mathbf{u}_\xi^{C_m,D_n} \otimes D_n \mathbf{v}_\xi^{C_m,D_n}$$

(cf. (4.8)), and performing the same type of downward and upward estimations as in Lemma 6.36 and Lemma 6.38 we conceive that

$$\|A\|_{*_{C_m,D_n}} = \sum_{\xi=1}^{\min(m,n)} \sigma_\xi^{C_m,D_n} = \sup \sum_{\psi=1}^{\min(m,n)} |\langle A\mathbf{y}_\psi, \mathbf{x}_\psi\rangle_{C_m}|,$$

where the supremum is taken over all orthonormal systems $(\mathbf{x}_\psi)_\psi$ and $(\mathbf{y}_\psi)_\psi$ in $(\mathbb{R}^m, \langle\,.\,,\,.\,\rangle_{C_m})$, respectively $(\mathbb{R}^n, \langle\,.\,,\,.\,\rangle_{D_n})$. On the other hand exploiting the assumed representation of A through the systems $(\lambda_\phi)_{\phi\in\{1,\ldots,\Phi\}}$, $(\mathbf{e}_\phi)_{\phi\in\{1,\ldots,\Phi\}}$ and $(\mathbf{f}_\phi)_{\phi\in\{1,\ldots,\Phi\}}$ the exact same argumentation leads to

$$\sup \sum_{\psi=1}^{\min(m,n)} |\langle A\mathbf{y}_\psi, \mathbf{x}_\psi\rangle_{C_m}| = \sum_{\phi=1}^{\Phi} \lambda_\phi$$

and the assertion is proven. □

With these insights we are now properly equipped to face the proof of Theorem 6.31.

Proof of Theorem 6.31.

(1) *Lim inf inequality:*
Let $(T_\mu)_{\mu\in\mathbb{N}} \subset \mathcal{I}^{2,2}(\Omega,\Sigma)$ be a convergent sequence with respect to the weak operator topology. Then, as already seen in the proof of Theorem 6.12, its limit T_* lies in $\mathcal{I}^{2,2}(\Omega,\Sigma)$ and it suffices to consider the scenario in which $\liminf_{\mu\to\infty} \mathcal{R}^{m_\mu,n_\mu}_{*_{C_{m_\mu},D_{n_\mu}}}(T_\mu)$ is bounded. Equivalently to the standard nuclear norm case this implies the existence of a weakly

convergent subsubsequence of $(T_\mu)_{\mu\in\mathbb{N}}$ which shares the same limit T_* and whose limit inferior with respect to $\mathcal{R}^{m_\mu,n_\mu}_{*C_{m_\mu},D_{n_\mu}}$ coincides with the one of the original sequence. Beyond that, this subsubsequence, which in the following will not be relabeled, satisfies

$$\sup_{\mu\in\mathbb{N}} \mathcal{R}^{m_\mu,n_\mu}_{*C_{m_\mu},D_{n_\mu}}(T_\mu) < \infty,$$

i.e. all of its elements can be represented by a matrix.
Now let $(\hat{e}_\phi)_{\phi\in\mathbb{N}}$ and $(\hat{f}_\phi)_{\phi\in\mathbb{N}}$ be arbitrary orthonormal systems in $L^2(\Sigma)$, respectively $L^2(\Omega)$, both with respect to the Euclidean inner product. Let furthermore $\Phi \in \mathbb{N}$. Then, applying Lemma 6.38 yields

$$\begin{aligned}\liminf_{\mu\to\infty} \mathcal{R}^{m_\mu,n_\mu}_{*C_{m_\mu},D_{n_\mu}}(T_\mu) &= \liminf_{\mu\to\infty} \sup_{\substack{(e_\psi)_\psi,(f_\psi)_\psi \\ \text{eucl. ONS}}} \sum_{\psi\in\mathbb{N}} \left| \left\langle T_\mu H_{D_{n_\mu}^{-\frac{1}{2}}} f_\psi, G_{C_{m_\mu}^{\frac{1}{2}}} e_\psi \right\rangle \right| \\ &\geq \liminf_{\mu\to\infty} \sum_{\phi=1}^{\Phi} \left| \left\langle T_\mu H_{D_{n_\mu}^{-\frac{1}{2}}} \hat{f}_\phi, G_{C_{m_\mu}^{\frac{1}{2}}} \hat{e}_\phi \right\rangle \right|\end{aligned}$$

and together with Fatou's Lemma we conceive that

$$\begin{aligned}\liminf_{\mu\to\infty} \mathcal{R}^{m_\mu,n_\mu}_{*C_{m_\mu},D_{n_\mu}}(T_\mu) \geq& \sum_{\phi=1}^{\Phi} \liminf_{\mu\to\infty} \left| \left\langle T_\mu H_{D_{n_\mu}^{-\frac{1}{2}}} \hat{f}_\phi, G_{C_{m_\mu}^{\frac{1}{2}}} \hat{e}_\phi \right\rangle \right| \\ \geq& \sum_{\phi=1}^{\Phi} \liminf_{\mu\to\infty} \left| \left\langle T_\mu D^{-\frac{1}{2}} \hat{f}_\phi, C^{\frac{1}{2}} \hat{e}_\phi \right\rangle \right| \\ &- \left| \left\langle T_\mu \left(D^{-\frac{1}{2}} - H_{D_{n_\mu}^{-\frac{1}{2}}} \right) \hat{f}_\phi, G_{C_{m_\mu}^{\frac{1}{2}}} \hat{e}_\phi \right\rangle \right| \\ &- \left| \left\langle T_\mu D^{-\frac{1}{2}} \hat{f}_\phi, \left(C^{\frac{1}{2}} - G_{C_{m_\mu}^{\frac{1}{2}}} \right) \hat{e}_\phi \right\rangle \right|.\end{aligned}$$

Elaborating on this latter expression we take a closer look on the subtrahends within the limit inferior and realize that for any fixed $\phi' \in \{1, \dots, \Phi\}$

$$\begin{aligned}\left| \left\langle T_\mu \left(D^{-\frac{1}{2}} - H_{D_{n_\mu}^{-\frac{1}{2}}} \right) \hat{f}_{\phi'}, G_{C_{m_\mu}^{\frac{1}{2}}} \hat{e}_{\phi'} \right\rangle \right| &\leq \|T_\mu\| \left\| \left(D^{-\frac{1}{2}} - H_{D_{n_\mu}}^{-\frac{1}{2}} \right) \hat{f}_{\phi'} \right\| \left\| G_{C_{m_\mu}^{\frac{1}{2}}} \right\|, \\ \left| \left\langle T_\mu D^{-\frac{1}{2}} \hat{f}_{\phi'}, \left(C^{\frac{1}{2}} - G_{C_{m_\mu}^{\frac{1}{2}}} \right) \hat{e}_{\phi'} \right\rangle \right| &\leq \|T_\mu\| \left\| D^{-\frac{1}{2}} \right\| \left\| \left(C^{\frac{1}{2}} - G_{C_{m_\mu}^{\frac{1}{2}}} \right) \hat{e}_{\phi'} \right\|\end{aligned}$$

holds true. Beyond that, respecting the pointwise convergence of $G_{C_{m_\mu}^{\frac{1}{2}}}$ and $H_{D_{n_\mu}^{-\frac{1}{2}}}$ toward $C^{\frac{1}{2}}$, respectively $D^{-\frac{1}{2}}$, we can even attest that for any $\varepsilon > 0$ there exists a ϕ'-dependent constant $\overline{M}_{\phi'} \in \mathbb{N}$ which ensures that

$$\left| \left\langle T_\mu \left(D^{-\frac{1}{2}} - H_{D_{n_\mu}^{-\frac{1}{2}}} \right) \hat{f}_{\phi'}, G_{C_{m_\mu}^{\frac{1}{2}}} \hat{e}_{\phi'} \right\rangle \right| \leq \varepsilon \|T_\mu\| \left\| G_{C_{m_\mu}^{\frac{1}{2}}} \right\|,$$
$$\left| \left\langle T_\mu D^{-\frac{1}{2}} \hat{f}_{\phi'}, \left(C^{\frac{1}{2}} - G_{C_{m_\mu}^{\frac{1}{2}}} \right) \hat{e}_{\phi'} \right\rangle \right| \leq \varepsilon \|T_\mu\| \left\| D^{-\frac{1}{2}} \right\|$$

for all $\mu \geq \overline{M}_{\phi'}$. Hence, involving the boundedness of $D^{-\frac{1}{2}}$ and $\sup_{\mu\in\mathbb{N}} ||T_\mu||$ as well as the boundedness of $\sup_{\mu\in\mathbb{N}} \|G_{C_{m_\mu}^{\frac{1}{2}}}\|$ which follows from the Banach-Steinhaus Theorem [Rudin, 1991, Thrm. 2.5], we are confident that both of these subtrahends become arbitrary small as soon as μ is big enough. Additionally incorporating that $(T_\mu)_{\mu\in\mathbb{N}}$ is weakly converging toward T_* we eventually perceive that

$$\liminf_{\mu\to\infty} \mathcal{R}^{m_\mu,n_\mu}_{*C_{m_\mu},D_{n_\mu}}(T_\mu) \geq \sum_{\phi=1}^{\Phi} | \left\langle T_* D^{-\frac{1}{2}} \hat{f}_\phi, C^{\frac{1}{2}} \hat{e}_\phi \right\rangle |.$$

Recalling that the orthonormal systems $(\hat{e}_\phi)_{\phi\in\mathbb{N}} \subset L^2(\Sigma)$ and $(\hat{f}_\phi)_{\phi\in\mathbb{N}} \subset L^2(\Omega)$ as well as the constant $\Phi \in \mathbb{N}$ were chosen arbitrarily, this implies together with Remark 6.37 the validity of

$$\begin{aligned}\liminf_{\mu\to\infty} \mathcal{R}^{m_\mu,n_\mu}_{*C_{m_\mu},D_{n_\mu}}(T_\mu) &\geq \sup_{\substack{(e_\phi)_\phi,(f_\phi)_\phi \\ \text{eucl. ONS}}} \sum_{\phi\in\mathbb{N}} | \left\langle T_* D^{-\frac{1}{2}} f_\phi, C^{\frac{1}{2}} e_\phi \right\rangle | \\ &= \mathcal{R}^{\infty}_{*C,D}(T_*).\end{aligned}$$

(2) *Existence of a recovery sequence:*

Within the proof of Theorem 6.12 we already saw that the set of integral operators with finite number of nonzero standard singular values lies dense in $\mathcal{I}^{2,2}(\Omega,\Sigma)$. In fact, exploiting the representation in Lemma 6.26 as well as the equivalence between the Euclidean inner product and $||\,.\,||_{[C,D^{-1}]}$ on $L^2(\Sigma\times\Omega)$ this insight can directly be extrapolated to the set of integral operators with finite number of nonzero C- and D-dependent singular values. Consequently, according to Remark 2.21 it suffices to prove the existence of a recovery sequence on this set.

So, let $T \in \mathcal{I}^{2,2}(\Omega,\Sigma)$ be an operator whose singular values $(\sigma_\xi^{C,D})_{\xi\in\mathbb{N}}$ regarding the inner products in $(L^2(\Sigma), \langle\,.\,,\,.\,\rangle_C)$ and $(L^2(\Omega), \langle\,.\,,\,.\,\rangle_D)$ equal zero as soon as $\xi > \Xi \in \mathbb{N}$. Denoting its associated singular system by $(\sigma_\xi^{C,D}, u_\xi^{C,D}, v_\xi^{C,D})_{\xi\in\mathbb{N}}$ this means that its corresponding integral kernel $t \in L^2(\Sigma\times\Omega)$ is of the form

$$t(s,r) = \sum_{\xi=1}^{\Xi} \sigma_\xi^{C,D} u_\xi^{C,D}(s) D v_\xi^{C,D}(r) \quad \forall s \in \Sigma, r \in \Omega$$

(cf. Lemma 6.26). Now reverting to the projections $P_m : L^2(\Sigma) \to L^2(\Sigma)$ and Q_n :

$L^2(\Omega) \to L^2(\Omega)$ introduced in (6.21) and assuming without loss of generality that μ is big enough to fulfill $\min(m_\mu, n_\mu) \geq \Xi$, we contemplate the systems $(P_{m_\mu} u_\xi^{C,D})_{\xi\in\{1,\dots,\Xi\}}$ and $(Q_{n_\mu} v_\xi^{C,D})_{\xi\in\{1,\dots,\Xi\}}$ and orthonormalize them following the construction in Definition 6.33. The resulting systems $(\bar{u}_\xi^\mu)_{\xi\in\{1,\dots,\Xi\}} \subset L^2(\Sigma)$ and $(\bar{v}_\xi^\mu)_{\xi\in\{1,\dots,\Xi\}} \subset L^2(\Omega)$ are then orthonormal with respect to the 'pseudo inner product' on piecewise constant functions induced by $G_{C_{m_\mu}}$, respectively $H_{D_{n_\mu}}$. With these systems at hand we define the sequence of operators $(T_\mu)_{\mu\in\mathbb{N}}$ via the sequence of its kernels $(t_\mu)_{\mu\in\mathbb{N}} \subset L^2(\Sigma\times\Omega)$ whose elements are characterized through

$$t_\mu(s,r) := \sum_{\xi=1}^{\Xi} \sigma_\xi^{C,D} \bar{u}_\xi^\mu(s) \left(H_{D_{n_\mu}} \bar{v}_\xi^\mu\right)(r).$$

Considering its elements discrepancy to the operator T we initially realize that

$$\begin{aligned}
\|T - T_\mu\|_{\mathcal{I}^{2,2}(\Omega,\Sigma)} &= \left\| \sum_{\xi=1}^{\Xi} \sigma_\xi^{C,D} u_\xi^{C,D} D v_\xi^{C,D} - \sum_{\xi=1}^{\Xi} \sigma_\xi^{C,D} \bar{u}_\xi^\mu \left(H_{D_{n_\mu}} \bar{v}_\xi^\mu\right) \right\|_{L^2(\Sigma\times\Omega)} \\
&\leq \sum_{\xi=1}^{\Xi} \sigma_\xi^{C,D} \left\| \left(u_\xi^{C,D} - \bar{u}_\xi^\mu\right) D v_\xi^{C,D} \right\|_{L^2(\Sigma\times\Omega)} \\
&\quad + \sum_{\xi=1}^{\Xi} \sigma_\xi^{C,D} \|\bar{u}_\xi^\mu \left(D - H_{D_{n_\mu}}\right) v_\xi^{C,D}\|_{L^2(\Sigma\times\Omega)} \\
&\quad + \sum_{\xi=1}^{\Xi} \sigma_\xi^{C,D} \left\|\bar{u}_\xi^\mu H_{D_{n_\mu}} \left(v_\xi^{C,D} - \bar{v}_\xi^\mu\right)\right\|_{L^2(\Sigma\times\Omega)} \\
&\leq \sum_{\xi=1}^{\Xi} \sigma_\xi^{C,D} \|D\| \left\|u_\xi^{C,D} - \bar{u}_\xi^\mu\right\|_{L^2(\Sigma)} \left\|v_\xi^{C,D}\right\|_{L^2(\Omega)} \\
&\quad + \sum_{\xi=1}^{\Xi} \sigma_\xi^{C,D} \|\bar{u}_\xi^\mu\|_{L^2(\Sigma)} \left\| \left(D - H_{D_{n_\mu}}\right) v_\xi^{C,D}\right\|_{L^2(\Omega)} \\
&\quad + \sum_{\xi=1}^{\Xi} \sigma_\xi^{C,D} \|\bar{u}_\xi^\mu\|_{L^2(\Sigma)} \left\|H_{D_{n_\mu}}\right\| \left\|v_\xi^{C,D} - \bar{v}_\xi^\mu\right\|_{L^2(\Omega)}
\end{aligned}$$

applies. Taking a closer look on the latter estimate we notice that, due to the pointwise convergence of $(H_{D_{n_\mu}})_{\mu\in\mathbb{N}}$ toward D and the in Lemma 6.34 demonstrated convergence of $(\bar{u}_\xi^\mu)_{\mu\in\mathbb{N}}$ and $(\bar{v}_\xi^\mu)_{\mu\in\mathbb{N}}$ to $u_\xi^{C,D}$, respectively $v_\xi^{C,D}$, all three subtractive expressions become arbitrary small as soon as μ is chosen big enough. Since in addition to that the Banach-Steinhaus Theorem [Rudin, 1991, Thrm. 2.5] ensures the boundedness of $\sup_{\mu\in\mathbb{N}} \|H_{D_{n_\mu}}\|$ and the convergence of $(\bar{u}_\xi^\mu)_{\mu\in\mathbb{N}}$ guarantees that $\sup_{\mu\in\mathbb{N}} \|\bar{u}_\xi^\mu\|$ is bounded we can infer that for any $\varepsilon > 0$ there is a global constant $M \in \mathbb{N}$ which certifies

$$\|T - T_\mu\|_{\mathcal{I}^{2,2}(\Omega,\Sigma)} \leq \varepsilon$$

for all $\mu \geq M$. Hence, $(T_\mu)_{\mu\in\mathbb{N}}$ converges (in norm) toward T.
With this convergence at hand we now want to verify that this candidate indeed represents a proper recovery sequence for T. In order to do so we first of all note, that by construction the elements in $(\bar{u}_\xi^\mu)_{\xi\in\{1,\dots,\Xi\}}$ and $(\bar{v}_\xi^\mu)_{\xi\in\{1,\dots,\Xi\}}$ are piecewise constant functions, i.e. there exist $(\mathbf{u}_\xi^\mu)_{\xi\in\{1,\dots,\Xi\}} \subset \mathbb{R}^{m_\mu}$ and $(\mathbf{v}_\xi^\mu)_{\xi\in\{1,\dots,\Xi\}} \subset \mathbb{R}^{n_\mu}$ with

$$\bar{u}_\xi^\mu = \sum_{i=1}^{m_\mu} (\mathbf{u}_\xi^\mu)_i \chi_{\Sigma_i^{m_\mu}}, \quad \bar{v}_\xi^\mu = \sum_{j=1}^{n_\mu} (\mathbf{v}_\xi^\mu)_j \chi_{\Omega_j^{n_\mu}}$$

for all $\xi \in \{1, \dots, \Xi\}$. Consequently, for all elements t_μ of our kernel sequence of choice it holds that

$$t_\mu = \sum_{i=1}^{m_\mu} \sum_{j=1}^{n_\mu} \left(\sum_{\xi=1}^{\Xi} \sigma_\xi^{C,D} (\mathbf{u}_\xi^\mu)_i \sum_{j'=1}^{n_\mu} |\Omega_{j'}^{n_\mu}|^{\frac{1}{2}} |\Omega_j^{n_\mu}|^{-\frac{1}{2}} (D_n)_{jj'} (\mathbf{v}_\xi^\mu)_{j'} \right) \chi_{\Sigma_i^{m_\mu}} \chi_{\Omega_j^{n_\mu}}$$

and we can deduce, that the corresponding elements T_μ are induced by the matrix

$$A_\mu := \left(\sum_{\xi=1}^{\Xi} \sigma_\xi^{C,D} (\mathbf{u}_\xi^\mu)_i \sum_{j'=1}^{n_\mu} |\Omega_{j'}^{n_\mu}|^{\frac{1}{2}} |\Omega_j^{n_\mu}|^{-\frac{1}{2}} (D_n)_{jj'} (\mathbf{v}_\xi^\mu)_{j'} \right)_{\substack{i=1,\dots,m_\mu \\ j=1,\dots,n_\mu}} \in \mathbb{R}^{m_\mu \times n_\mu},$$

i.e. $T_\mu = EA_\mu$ for all $\mu \in \mathbb{N}$. Contemplating the associated matrix $(\omega \cdot A_\mu)$ which is consulted when computing the nuclear norm of T_μ, we then observe that for systems $(\hat{\mathbf{u}}_\xi^\mu)_{\xi\in\{1,\dots,\Xi\}} \subset \mathbb{R}^{m_\mu}$ and $(\hat{\mathbf{v}}_\xi^\mu)_{\xi\in\{1,\dots,\Xi\}} \subset \mathbb{R}^{n_\mu}$ whose elements are defined via

$$\hat{\mathbf{u}}_\xi^\mu := \left(|\Sigma_i^{m_\mu}|^{\frac{1}{2}} (\mathbf{u}_\xi^\mu)_i \right)_{i=1,\dots,m_\mu}, \quad \hat{\mathbf{v}}_\xi^\mu := \left(|\Omega_j^{n_\mu}|^{\frac{1}{2}} (\mathbf{v}_\xi^\mu)_j \right)_{j=1,\dots,n_\mu}$$

the following identity applies:

$$\omega \cdot A_\mu = \left(\sum_{\xi=1}^{\Xi} \sigma_\xi^{C,D} (\hat{\mathbf{u}}_\xi^\mu)_i (D_n \hat{\mathbf{v}}_\xi^\mu)_j \right)_{\substack{i=1,\dots,m_\mu \\ j=1,\dots,n_\mu}}.$$

However, recalling that in Remark 6.32 we already found that

$$\langle \bar{u}_\xi^\mu, G_{C_{m_\mu}} \bar{u}_\psi^\mu \rangle_{L^2(\Sigma)} = \langle \hat{\mathbf{u}}_\xi^\mu, C_{m_\mu} \hat{\mathbf{u}}_\psi^\mu \rangle_{\mathbb{R}^{m_\mu}}, \quad \langle \bar{v}_\xi^\mu, H_{D_{n_\mu}} \bar{v}_\psi^\mu \rangle_{L^2(\Omega)} = \langle \hat{\mathbf{v}}_\xi^\mu, D_{n_\mu} \hat{\mathbf{v}}_\psi^\mu \rangle_{\mathbb{R}^{n_\mu}}$$

and taking into account that $(\bar{u}_\xi^\mu)_{\xi\in\{1,\dots,\Xi\}}$ and $(\bar{v}_\xi^\mu)_{\xi\in\{1,\dots,\Xi\}}$ were constructed to be orthonormal with respect to the 'pseudo inner products' induced by $G_{C_{m_\mu}}$ and $H_{D_{n_\mu}}$ we realize that $(\hat{\mathbf{u}}_\xi^\mu)_{\xi\in\{1,\dots,\Xi\}}$ and $(\hat{\mathbf{v}}_\xi^\mu)_{\xi\in\{1,\dots,\Xi\}}$ are orthonormal with respect to $\langle\, . \,, . \,\rangle_{C_{m_\mu}}$, respectively $\langle\, . \,, . \,\rangle_{D_{n_\mu}}$. Therefore, incorporating the statement in Lemma 6.39 we deduce that

$$\mathcal{R}^{m_\mu,n_\mu}_{*C_{m_\mu},D_{n_\mu}}(T_\mu) = \sum_{\eta=1}^{\min(m_\mu,n_\mu)} \sigma_\eta^{C_{m_\mu},D_{n_\mu}}(\omega \cdot A_\mu) = \sum_{\xi=1}^{\Xi} \sigma_\xi^{C,D}$$

as $\min(m_\mu, n_\mu) \geq \Xi$ and eventually obtain

$$\lim_{\mu\to\infty} \mathcal{R}^{m_\mu,n_\mu}_{*C_{m_\mu},D_{n_\mu}}(T_\mu) = \sum_{\xi=1}^{\Xi} \sigma_\xi^{C,D} = \mathcal{R}^{\infty}_{*C,D}(T).$$

□

With this proof of Γ-convergence we now completed the asymptotic analysis of functionals of the form (6.7) which involve the generalized nuclear norm as a regularizer. Combined with the results in Proposition 6.5 and 6.7 we hence derived a limit functional which due to its preserving minimizing structure can be consulted when dealing with very high-dimensional variational problems.

7 Conclusion and outlook

In this thesis, we discussed how the regularization with nuclear norms can support the variational reconstruction of dynamic MR scans.
After a brief introduction to the concept of variational methods, in Chapter 4 we first demonstrated how the singular value decomposition of matrices, whose columns represent the vectorized frames of a time series, decomposes them into their individual pairwise orthogonal dynamics. In doing so, we figured out, that the spatial localization of the observed dynamics as well as their temporal evolution is captured in the right and left singular vectors. Furthermore, we found that via the associated singular values also a certain dominance is assigned to these artifacts: While dynamics with a large singular value are considered to have a big influence on the course of the time series, those with a smaller singular value are perceived as less influential.
With the help of this understanding, we then turned to modeling explicit minimization problems for the reconstruction of dynamic MR data. To this end, we first reviewed some existing approaches that, with respect to regularization, rely on the nuclear norm. In scenarios where the MR data is assumed to encode only few observable dynamics, good results can be obtained via this family of approaches, since the nuclear norm coincides with the ℓ^1-norm on the singular values and therefore promotes their sparsity. In terms of our interpretative findings, this means that only time series that can be decomposed into a few influential dynamics can be considered as solutions to such minimization problems.
In our review, we also focused on the low rank + sparse approach, which represents the sought-after reconstruction as the superposition of two individual time series, only one of which is penalized using the nuclear norm. With respect to the second time series, the model is supplemented by the regularization with the matrix $1,1$-norm. This promises a particularly good result in scenarios where the reconstruction is expected to identify the motion of very small structures against an almost constant background.
In a second step, we then dealt with expanding the regularization with the nuclear norm. Guided by our previously gained interpretative understanding, we set out to influence the perception of dominance of dynamics induced via the singular value decomposition. For this purpose, we introduced a generalized form of the classical SVD: Perceiving matrices as linear mappings between non-euclidean vector spaces allowed us to manipulate the singular vectors as well as the singular values by only choosing appropriate positive definite and symmetric matrices. In particular, we encountered interesting results when considering an approximation of the negative Laplacian matrix. Namely, in this special case we found that the singular values to singular

vectors with large discretized H^1-norm turn out rather small. Conversely, large singular values were assigned to those singular vectors that are associated with particularly smooth dynamics. This observation led us to also define a generalized nuclear norm and to incorporate its concrete Laplacian expression as a regularizer in the minimization problem of reconstruction. Following our findings from the classical case, we expected this model to promote the occurrence of smooth dynamics and thus to support the reconstruction of those dynamical data from which such behavior is expected.
In Chapter 5, we then addressed to verify the effectiveness of the introduced variational models by using explicit application-based examples. To this end, we first of all derived concrete iterative schemes for their numerical solution based on the forward-backward splitting algorithm. With these we were able to demonstrate the effectiveness of the low rank + sparse approach on simulated raw MR data of a brain. Despite heavy undersampling, this method provided a reconstruction that reproduced even the movements of very small cells. Moreover, using the real MR scan of a mouse brain, we showed that a variation of this approach is also suitable to perform posterior cell tracking in already reconstructed dynamic series. Regarding the regularization with the newly introduced generalized nuclear norm, we first wanted to approve numerically that, analogous to the classical nuclear norm, in combination with the identity operator its incorporation can indeed be traced back to a linear shrinkage of the singular values. Surprisingly, however, we found that the generalized singular values in the previously introduced Laplacian setting do not undergo uniform diminution. Instead, we observed a strong signal-dependent behavior. This makes it currently unpredictable how the dominance of singular vectors will be perceived after the minimization problem is solved. However, since the variational problem considered here can be reformulated to match the one which involves a real (smoothing) operator and regularizes with the classical nuclear norm, with this observation we were also able to point out that even the more established model requires further investigation.
Within Chapter 6 we then took care of the asymptotic analysis of the functionals considered in the applied part of this work. Concentrating on their behavior as temporal and/or spatial dimensions tend to infinity, we first of all showed the general equi-coercivity of this family of problems with respect to the weak operator topology on the space of (semi-discrete) integral operators. Subsequently, we found that the involved data fidelity term converges continuously toward its natural continuous counterpart. Hence, aiming for Γ-convergence, in the following we focused on the determination of the respective Γ-limits of the included regularizing norms. In doing so, we were ultimately able to confirm that the mixed p, q-norm for $p, q > 1$, as well as the classical and generalized nuclear norm Γ-converge toward their logical continuous counterparts in the space of $L^{p,q}$-integral operators, respectively $L^{2,2}$-integral operators. Altogether, we have thus successfully transferred the discretely defined minimization problems into a continuous setting, and therefore provided the basis for further more analytical investigations.

Although we have gained some new insights in this work, we certainly did not tackle all open questions on the reconstruction of dynamic MR scans via the regularization with the nuclear norm. Instead, by introducing the generalized nuclear norm we have raised many new ones. First and foremost, it would be desirable to understand the pattern in which the generalized singular values are shrunk during optimization with the identity operator. As mentioned before, this question can be traced back to minimization problems with true forward operator and regularizing classical nuclear norm. Hence, in a first step, future work on this topic should definitely address the general interaction between involved operators and the associated shrinkage behavior of singular values. Based on these results, the study of regularization with the generalized nuclear norm can then be resumed in the context of dynamic MRI.

In addition, of course, it is worth considering in what other contexts this new norm might be useful. Here it is used to favor smooth signals in dynamic series, however, that is probably just one of many potential applications. By choosing wisely the positive definite matrices on which it depends, the generalized nuclear norm offers a large number of possibilities to integrate additional knowledge in the solution of inverse problems.

But also regarding the theoretical part of this thesis there is an open point left. In fact, in order to prove the Γ-convergence of the generalized nuclear norm, we made some assumptions which ultimately led to the considered modified norms being equivalent to the Euclidean norm. This simplified the analysis considerably for us. We thus leave the proof without these presumptions to future work.

All in all, we hope that with this work we have been able to draw general attention to the versatility of the nuclear norm. Although its use as a regularizer for solving inverse problems has been rather underrepresented so far, it is worth thinking about more far-reaching applications. With the definition of its generalization, we hope to contribute to expand its radius of impact.

Bibliography

ACM SIGKDD and Netflix (2007). Proceedings of KDD Cup and Workshop. Proceedings available online at http://www.cs.uic.edu/ liub/KDD-cup-2007/proceedings.html.

Bauschke, H. H. and Combettes, P. L. (2011). *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer.

Benedek, A. and Panzone, R. (1961). The space L^p with mixed norm. *Duke Mathematical Journal*, 28(3):301–324.

Boyer, C., De Castro, Y., and Salmon, J. (2017). Adapting to unknown noise level in sparse deconvolution. *Information and Inference: A Journal of the IMA*, 6(3):310–348.

Braides, A. (2002). Γ-*convergence for Beginners*. Oxford lecture series in mathematics and its applications. Oxford University Press.

Bredies, K. and Pikkarainen, H. K. (2013). Inverse problems in spaces of measures. *ESAIM: Control, Optimisation and Calculus of Variations*, 19(1):190–218.

Brezis, H. (2010). *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. Springer New York.

Brinkmann, E.-M. (2019). *Novel Aspects of Total Variation-Type Regularization in Imaging*. PhD thesis, Westfälische Wilhelms-Universität Münster.

Cai, J.-F., Candès, E. J., and Shen, Z. (2010). A Singular Value Thresholding Algorithm for Matrix Completion. *SIAM Journal on Optimization*, 20(4):1956–1982.

Candès, E. J., Li, X., Ma, Y., and Wright, J. (2011). Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):1–37.

Candès, E. J. and Recht, B. (2009). Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772.

Candès, E. J., Romberg, J., and Tao, T. (2006a). Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509.

Candès, E. J., Romberg, J., and Tao, T. (2006b). Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223.

Candès, E. J. and Tao, T. (2005). Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215.

Candès, E. J. and Tao, T. (2010). The Power of Convex Relaxation: Near-Optimal Matrix Completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080.

Cauchy, A. (1847). Méthode générale pour la résolution des systemes d'équations simultanées. *Comptes Rendus de l'Académie des Sciences*, 25:536–538.

Chambolle, A., De Vore, R. A., Lee, N.-Y., and Lucier, B. J. (1998). Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Transactions on Image Processing*, 7(3):319–335.

Chambolle, A. and Pock, T. (2016). An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319.

Chandrasekaran, V., Sanghavi, S., Parillo, P., and Willsky, A. (2011). Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596.

Cocosco, C. A., Kollokian, V., Kwan, R. K.-S., and Evans, A. C. (1997). BrainWeb: Online Interface to a 3D MRI Simulated Brain Database. In *NeuroImage*. Citeseer.

Cohen, A., Dahmen, W., and Devore, R. (2009). Compressed sensing and best k-term approximation. *Journal of the American Mathematical Society*, 22(1):211–231.

Colton, D., Engl, H. W., Louis, A. K., McLaughlin, J., and Rundell, W. (2012). *Surveys on solution methods for inverse problems*. Springer Science & Business Media.

Combettes, P. L. and Wajs, V. R. (2005). Signal Recovery by Proximal Forward-Backward Splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200.

Dal Maso, G. (1993). *An Introduction to* Γ*-Convergence*. Progress in Nonlinear Differential Equations and Their Applications. Birkhäuser Boston.

Dashti, M., Law, K. J. H., Stuart, A. M., and Voss, J. (2013). MAP Estimators and their Consistency in Bayesian Nonparametric Inverse Problems. *Inverse Problems*, 29(9).

Daubechies, I., Defrise, M., and Mol, C. D. (2004). An iterative thresholding algorithm for linear inverseproblems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457.

Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306.

Donoho, D. L. and Elad, M. (2003). Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202.

Donoho, D. L. and Huo, X. (2001). Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47(7):2845–2862.

Dunford, N. and Schwartz, J. T. (1988). *Linear Operators, Part 1: General Theory*, volume 10. John Wiley & Sons.

Ekeland, I. and Temam, R. (1999). *Convex Analysis and Variational Problems*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics.

Elster, A. D. and Burdette, J. H. (2001). *Questions & Answers in Magnetic Resonance Imaging*. Mosby.

Engl, H. W. (1997). *Integralgleichungen*. Springer-Lehrbuch: Mathematik. Springer Vienna.

Engl, H. W., Hanke, M., and Neubauer, A. (1996). *Regularization of Inverse Problems*. Mathematics and Its Applications. Springer Netherlands.

Fan, K. (1951). Maximum Properties and Inequalities for the Eigenvalues of Completely Continuous Operators. *Proceedings of the National Academy of Sciences of the United States of America*, 37(11):760–766.

Fazel, M., Hindi, H., and Boyd, S. (2001). A rank minimization heuristic with application to minimum order system approximation. In *Proceedings American Control Conference*, pages 4734–4739.

Figueiredo, M. A. T. and Nowak, R. D. (2003). An EM algorithm for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 12(8):906–916.

Fornasier, M. (2010). *Theoretical Foundations and Numerical Methods for Sparse Recovery*. Radon Series on Computational and Applied Mathematics. De Gruyter.

Gao, H., Rapacchi, S., Wang, D., Moriarty, J., Meehan, C., Sayre, J., Laub, G., Finn, P., and Hu, P. (2012). Compressed Sensing using Prior Rank, Intensity and Sparsity Model (PRISM): Applications in Cardiac Cine MRI. In *Proceedings of the 20th Annual Meeting of ISMRM, Melbourne, Australia*, page 2242.

Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Groetsch, C. W. (1993). *Inverse problems in the mathematical sciences*, volume 52. Springer.

Hadamard, J. (1902). Sur les Problèmes aux Dérivées Partielles et Leur Signification Physique. *Princeton University Bulletin*, 13:49–52.

Haldar, J. and Liang, Z.-P. (2010). Spatiotemporal imaging with partially separable functions: a matrix recovery approach. In *Proceedings IEEE International Symposium of Biomedical Imaging*, pages 716–719.

Heins, P. (2014). *Reconstruction Using Local Sparsity: A Novel Regularization Technique and an Asymptotic Analysis of Spatial Sparsity Priors*. PhD thesis, Westfälische Wilhelms-Universität Münster.

Helin, T. and Burger, M. (2015). Maximum a Posteriori Probability Estimates in Infinite-Dimensional Bayesian Inverse Problems. *Inverse Problems*, 31(8).

Hemmer, B., Kerschensteiner, M., and Korn, T. (2015). Role of the innate and adaptive immune responses in the course of multiple sclerosis. *The Lancet Neurology*, 14(4):406–419.

Horn, R. A. and Johnson, C. R. (1994). *Topics in Matrix Analysis*. Cambridge University Press.

Huettel, S. A., Song, A. W., and McCarthy, G. (2004). *Functional magnetic resonance imaging*. Sinauer Associates Sunderland, MA.

Hytönen, T., van Neerven, J., Veraar, M., and Weis, L. (2016). *Analysis in Banach Spaces - Volume I: Martingales and Littlewood-Paley Theory*, volume 12. Springer Berlin.

Jackson, J. I., Meyer, C. H., Nishimura, D. G., and Macovski, A. (1991). Selection of a convolution function for fourier inversion using gridding (computerised tomography application). *IEEE Transactions on Medical Imaging*, 10(3):473–478.

Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer Series in Statistics. Springer.

Kirsch, A. (2011). *An Introduction to the Mathematical Theory of Inverse Problems*. Applied Mathematical Sciences. Springer New York.

Lauterbur, P. (1973). Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance. *Nature*, 242:190–191.

Lebedev, L. P., Vorovich, I. I., and Gladwell, G. M. L. (2003). *Functional Analysis - Applications in Mechanics and Inverse Problems*. Solid Mechanics and Its Applications. Kluwer Academic Publishers.

Lewis, A. S. (2003). The mathematics of eigenvalue optimization. *Mathematical Programming*, 97(1):155–176.

Liang, Z.-P. (2007). Spatiotemporal imaging with partially separable functions. In *Proceedings IEEE International Symposium of Biomedical Imaging*, pages 181–182.

Liang, Z.-P. and Lauterbur, P. C. (2000). *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective*. IEEE Press series in biomedical engineering. SPIE Optical Engineering Press.

Lingala, S. G., Hu, Y., DiBella, E., and Jacob, M. (2011). Accelerated dynamic MRI exploiting sparsity and low-rank structure: kt SLR. *IEEE transactions on medical imaging*, 30(5):1042–1054.

Lions, P. L. and Mercier, B. (1979). Splitting Algorithms for the Sum of Two Nonlinear Operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979.

Lustig, M., Donoho, D., and Pauly, J. (2007). Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic resonance in medicine*, 58(6):1182–1195.

Masthoff, M., Freppon, F., Wachsmuth, L., Albers, F., Schwarz, C., Gerwing, M., Helfen, A., Heindel, W., Eisenblaetter, M., Wildgruber, M., and Faber, C. (2019). Time-lapse MRT als dynamische single cell tracking Bildgebungstechnik zur Entzündungsdiagnostik. In *Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*, volume 191.

Megginson, R. E. (2012). *An Introduction to Banach Space Theory*. Graduate Texts in Mathematics. Springer New York.

Modersitzki, J. (2009). *FAIR: Flexible Algorithms for Image Registration*. SIAM.

Moreau, J. (1965). Proximité et dualité dans un espace hilbertien. *Bulletin de la société mathématique de France*, 93:273–299.

Noll, D., Cohen, J., Meyer, C., and Schneider, W. (1995). Spiral k-space MR imaging of cortical activation. *Journal of magnetic resonance imaging: JMRI*, 5(1):49–56.

O'Sullivan, J. (1985). A Fast Sinc Function Gridding Algorithm for Fourier Inversion in Computer Tomography. *IEEE Transactions on Medical Imaging*, 4(4):200–207.

Otazo, R., Candès, E., and Sodickson, D. (2015). Low-Rank Plus Sparse Matrix Decomposition for Accelerated Dynamic MRI with Separation of Background and Dynamic Components. *Magnetic Resonance in Medicine*, 73(3):1125–1136.

Petrella, J. R. and Provenzale, J. M. (2000). MR perfusion imaging of the brain: techniques and applications. *American Journal of roentgenology*, 175(1):207–219.

Rasch, J. (2018). *Advanced Convex Analysis for Improved Variational Image Reconstruction*. PhD thesis, Westfälische Wilhelms-Universität Münster.

Recht, B., Fazel, M., and Parrilo, P. (2010). Guaranteed minimum rank solutions of matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501.

Riesz, F. and Sz-Nagy, B. (1955). *Functional Analysis*. F. Ungar Publishing Company.

Rudin, W. (1991). *Functional Analysis*. International series in pure and applied mathematics. McGraw-Hill.

Samarah, S., Obeidat, S., and Salman, R. (2005). A Schur test for weighted mixed-norm spaces. *Analysis Mathematica*, 31(4):277–289.

Sánchez, A., Schimmang, T., and García-Sancho, J. (2012). *Cell and Tissue Therapy in Regenerative Medicine*, pages 89–102. Springer New York.

Schatten, R. (2013). *Norm Ideals of Completely Continuous Operators*, volume 27. Springer Berlin.

Siemens Healthcare GmbH (2015). *Magnets, Spins, and Resonances: An Introduction to the Basics of Magnetic Resonance*. Manuscript, Siemens Healthcare GmbH.

Stuart, A. M. (2010). Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society*, 58(1):267–288.

Tikhonov, A. N. and Arsenin, V. I. A. (1977). *Solutions of ill-posed problems*. Scripta series in mathematics. Winston.

Tikhonov, A. N., Goncharsky, A., and Bloch, M. (1987). *Ill-posed problems in the natural sciences*. Mir Moscow.

Tropp, J. A. (2004). Greed is good: algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10):2231–2242.

Watson, G. A. (1992). Characterization of the subdifferential of some matrix norms. *Linear Algebra and its Applications*, 170:1039–1053.

Wernick, M. N. and Aarsvold, J. (2004). *Emission Tomography: The Fundamentals of PET and SPECT*. Elsevier Science.

Xiang, Q. S. and Henkelman, R. M. (1993). K-space description for MR imaging of dynamic objects. *Magnetic resonance in medicine*, 29(3):422–428.

Yue, M. and So, A. (2016). A Perturbation Inequality for the Schatten-p Quasi-Norm and Its Applications to Low-Rank Matrix Recovery. *Applied and Computational Harmonic Analysis*, 40(2):396–416.

Zhao, B., Haldar, J. P., Christodoulou, A. G., and Liang, Z.-P. (2012). Image reconstruction from highly undersampled (k, t)-space data with joint partial separability and sparsity constraints. *IEEE transactions on medical imaging*, 31(9):1809–1820.

List of figures

4.1 Spin ensemble during the exposition to an external magnetic field 34
4.2 Spin ensemble before and at the end of a 90-degree pulse 34
4.3 The singular value decomposition of a spatio-temporal matrix 46
4.4 The Laplacian-based generalized singular vectors of a spatio-temporal matrix . 52

5.1 Reconstruction results emerging from the low rank + sparse approach 64
5.2 Cell tracking results emerging from a variation of the low rank + sparse approach 66
5.3 Classical and generalized singular vectors after solving the denoising variational problem with regularizing classical or generalized nuclear norm and varying regularization parameter 69
5.4 Classical and generalized singular values after solving the denoising variational problem with regularizing classical or generalized nuclear norm and varying regularization parameter 69

List of tables

4.1 The Laplacian-based generalized singular values of a spatio-temporal matrix . . 53