Regularization Graphs — A unified framework for variational regularization of inverse problems

Kristian Bredies[*]   Marcello Carioni[†]   Martin Holler [‡]

Abstract

We introduce and study a mathematical framework for a broad class of regularization functionals for ill-posed inverse problems: *Regularization Graphs*. Regularization graphs allow to construct functionals using as building blocks linear operators and convex functionals, assembled by means of operators that can be seen as generalizations of classical infimal convolution operators. This class of functionals exhaustively covers existing regularization approaches and it is flexible enough to craft new ones in a simple and constructive way. We provide well-posedness and convergence results with the proposed class of functionals in a general setting. Further, we consider a bilevel optimization approach to learn optimal weights for such regularization graphs from training data. We demonstrate that this approach is capable of optimizing the structure and the complexity of a *regularization graph*, allowing, for example, to automatically select a combination of regularizers that is optimal for given training data.

Key words: Inverse problem, regularization functional, graph data structure, bilevel optimization.

2020 Mathematics Subject Classification: 65J20, 65K10, 49J45.

1   Introduction

In the last decades, a significant part of inverse problems theory has revolved around constructing suitable regularization approaches that allow for a reliable solution of ill-posed inverse problems. Among those, energy-based methods such as Tikhonov regularization [50] have been successful both with respect to mathematical guarantees, e.g., on well-posedness and stability, and with respect to practical performance in applications. An important cornerstone of energy-based methods are regularization functionals, which are responsible for stabilizing the ill-posed inversion of the forward model and for incorporating prior knowledge, such as smoothness, on the sought solution. The later is relevant in particular when dealing with highly structured data such as image data, where a suitable inclusion of prior knowledge makes a significant difference regarding the overall performance of the resulting method.

[*]Kristian Bredies, Institute of Mathematics and Scientific Computing, University of Graz, Heinrichstraße 36, A-8010 Graz, Austria. Email: `kristian.bredies@uni-graz.at`

[†]Marcello Carioni, University of Cambridge, Department of Applied Mathematics and Theoretical Physics, Wilberforce Road, Cambridge CB3 0WA, UK Email: `mc2250@maths.cam.ac.uk`

[‡]Martin Holler, Institute of Mathematics and Scientific Computing, University of Graz, Heinrichstraße 36, A-8010 Graz, Austria. Email: `martin.holler@uni-graz.at`
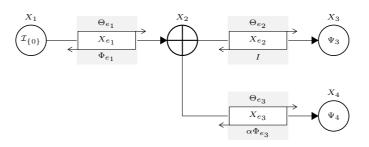
Figure 1: Example of a *regularization graph* with nodes $V = \{1, 2, 3, 4\}$, edges $E = \{(1,2), (2,3), (2,4)\} = \{e_1, e_2, e_3\}$ and weights $(\alpha_{(1,2)}, \alpha_{(2,3)}, \alpha_{(2,4)}) = (1, 1, \alpha)$. See Remark 2.5 for a detailed interpretation.

In this context, non-smooth sparsity-based methods building on measures, measure-valued differential operators, basis transforms or frames have become very popular. Besides the celebrated total variation (TV) functional [46], those include methods building on higher-order derivatives such as second-order TV [31], infimal-convolution-based approaches [18] or the total generalized variation (TGV) functional [8], see [7] for a recent review. Transform-based methods include wavelet-, curvelet- or shearlet transforms [36, 39, 49] as well as learned dictionaries [23].

Also more specific approaches tailored, for instance, to model certain oscillations [25, 27, 33, 38, 40, 43, 44] or texture [17], as well as different combinations of existing methods exist, such as TV and second-order TV [42], higher-order regularizers [12, 19, 42], TV-type functionals with curvelets or shearlets [26, 30, 28], a combination of different transform-based approaches [35] or the infimal convolution of TV with L^p-norms [14, 15]. We refer to [7, 3, 9, 37, 47] for a review of a subset of the plethora of existing methods.

While all these approaches share the goal of providing a model-based regularization for inverse problems, the way and extent to which they are developed and analyzed is rather different and often application-specific. Moreover, the choice of any of such methods is mostly done manually. A systematic approach for the analysis and the automatic, data-based design of regularization functionals that covers a broad class of existing methods does not exist to date. With introducing the framework of regularization graphs, we aim to provide a step in this direction. A regularization graph can be described as a weighted, directed graph together with a collection of functionals and operators associated with the nodes and the edges of the graph, respectively. Such structure allows to define regularization functionals via a rather arbitrary combination of linear operators and functionals, e.g., via variable splittings or summations. In particular, both the sum and a (generalized) infimal convolution of the functionals associated with two regularization graphs can be formulated as a regularization graph functional, where the underlying graph is obtained by properly combining the two original ones.

This yields a flexible framework for designing new regularization functionals or combining existing ones, e.g., via infimal convolution. Moreover, by associating weights to the edges of such graphs, a learning of both the parameters associated with such functionals as well as the structure of the underlying graph is possible. The latter in particular allows to automatically select optimal regularization functionals from a set of possible choices within a bilevel approach.

A prototypical example of a regularization graph with nodes $V = \{1, 2, 3, 4\}$, directed edges $E = \{(1,2), (2,3), (2,4)\}$ and weights $(\alpha_e)_{e \in E}$ is provided in Figure 1. Here, the operators and

the functionals associated to the nodes and the edges of the graph are defined as follows. For $n \in V$ and $e \in E$, X_n and X_e are suitable Banach spaces, Φ_e are bounded linear operators, Θ_e are (possibly unbounded) closed range operators and Ψ_n are convex functionals. The spaces X_n and X_e are called *node spaces* and *edge spaces*, respectively, while the Ψ_n are called *node functionals*. Further, we call Θ_e *forward operators* as they map from the edge space X_e to the direct successor node space X_n. Similarly, we call Φ_e *backward operators* as they map from the edge space to the direct predecessor node space. Variables $\{w_{e_1}, w_{e_2}, w_{e_3}\}$ associated with the edges of the graph, on which both the forward and backward operators are evaluated, are called *edge variables*.

Notice that in our example the root node 1 and the splitting node 2 correspond to the functional $\mathcal{I}_{\{0\}}$, i.e., $\Psi_1 = \Psi_2 = \mathcal{I}_{\{0\}}$ and $\Phi_{e_2} = I$ denotes a continuous embedding of X_{e_2} into X_2; see Remark 2.5 below for details. Also note that the weights $(\alpha_e)_{e \in E}$ associated to the graph are depicted in Figure 1 as scalar factors in front of the backward operators $(\Phi_e)_{e \in E}$, where we use the convention that fixed, trivial weights $\alpha_e = 1$ are not depicted explicitly. We also remark that, besides the notation α_e for $e \in E$, for specific regularization graphs the non-trivial weights will be often numbered independently of the edge they are associated with; see Figure 2.

The regularization functional $R_\alpha : X_1 \to [0, +\infty]$ corresponding to such a regularization graph is given via minimizing over the involved edge variables $\{w_{e_1}, w_{e_2}, w_{e_3}\}$ as

$$R_\alpha(u) = \inf_{(w_{e_i})_i} \mathcal{I}_{\{0\}}(u - \Phi_{e_1} w_{e_1}) + \mathcal{I}_{\{0\}}(\Theta_{e_1} w_{e_1} - w_{e_2} - \alpha \Phi_{e_3} w_{e_3}) + \Psi_3(\Theta_{e_2} w_{e_2}) + \Psi_4(\Theta_{e_3} w_{e_3})$$

$$= \inf_{w_{e_1}, w_{e_3}} \Psi_3(\Theta_{e_2}(\Theta_{e_1} w_{e_1} - \alpha \Phi_{e_3} w_{e_3})) + \Psi_4(\Theta_{e_3} w_{e_3}) \quad \text{s.t.} \ \ u = \Phi_{e_1} w_{e_1}.$$

The structure defined in this example is a regularization graph under mild additional conditions, most importantly weak* lower semicontinuity and coercivity of Ψ_3 and Ψ_4, and closedness of the range of each Θ_e, which, for instance, still allows the Θ_e to be densely defined differential operators and the Φ_e to be synthesis operators for a given dictionary or frame. The non-trivial weight α allows to adapt the structure of the graph by removing edges, as with $\alpha = 0$ and supposing for example that Ψ_4 vanishes in zero, we obtain $R_0(u) = \inf_{\{\Phi_{e_1} w_{e_1} = u\}} \Psi_3(\Theta_{e_2} \Theta_{e_1} w_{e_1})$.

The general structure of a regularization graph is defined in Section 2 and examples of existing regularization approaches that are included in this setting are provided in Section 2.1 and listed in the Appendix. Here, the main conditions on the involved functionals and operators are that the forward operators Θ_e have closed range (i.e., satisfy a Poincaré-type estimate), that the backward operators Φ_e are continuous and that the involved node functionals Ψ_n are coercive.

Under these conditions, we prove well-posedness, stability and convergence results for the application of regularization graphs in a general inverse problem setting. Moreover, we develop a bilevel approach that allows to learn the structure of an optimal graph for a given set of training data and show well-posedness of the resulting non-convex optimization problem.

Contribution of the paper in relation to the state of the art. In a rather abstract setting, general conditions on regularization functionals that allow to guarantee well-posedness, stability and convergence are of course well-known, see for instance [29, 32]. Those, however, are conditions on the overall functionals rather than their building blocks and their verification is often at the same level of difficulty than the results themselves. Furthermore, they do

3

not allow to easily combine different approaches without re-checking the underlying conditions. More specific results also exist, but deal with particular settings such as higher-order regularization [6, 12].

More related to the aim of this paper are some works on bilevel optimization, see for instance [16] for a review. In the probably most closely related work [21], the authors consider a general bilevel framework that includes TV, the infimal convolution of first and second order TV functionals as well as the TGV functional as particular cases. In contrast to [21], however, where essentially well-posed linear inverse problems are considered, i.e., those with closed range forward operator, our work is generally applicable to any bounded forward operator. In particular, we do not require closed range and allow for genuinely ill-posed inverse problems, a generalization that is the main source of difficulty for the analysis in this context.

A second, closely related work is the preprint [20]. There, the authors consider a bilevel scheme for learning parameters and operators in a TGV-like functional. They provide conditions on the involved operators under which they show well-posedness for a bilevel approach in image denoising. As application they consider an interpolation between a symmetrized and a non-symmetrized differential operator in the second order TGV functional. Besides being applicable to inverse problems beyond denoising, our work is different to [20] in allowing a more flexible combination of linear operators and functionals, far beyond the cascadic structure of TGV. Further, our framework allows for an automatic selection from different choices of existing regularization functionals but also, for instance, to select an optimal order in TGV regularization.

Organization of the paper. The paper is organized as follows. In Section 2 we give the precise definition of *regularization graphs* clarifying the main assumptions on the linear operators, the functionals and the involved Banach spaces that yield the results of our work. Also, we provide several examples of existing regularization approaches that can be constructed using a suitable *regularization graph*. In Section 3 we provide basic algebraic properties of regularization graphs, in particular a recursive representation that will be quite useful later on. In Section 4 we provide the main analytic properties of functionals associated with regularization graphs that will be the basis for subsequent results on the regularization of inverse problems and bilevel optimization. In particular, we show that any such functional is weak* lower semi-continuous and coercive up to a finite dimensional space. In Section 5 we provide an equivalent predual formulation of regularization graphs. Also, the connection to well-known predual representations of existing regularization approaches is made. While the results of this section will not be needed in the subsequent theory, they are nevertheless of interest on their own, in particular in view of optimality conditions and duality-based algorithms.

Section 6 then provides well-posedess and convergence results for the application of regularization graphs to the regularization of linear inverse problems. We focus on linear inverse problems since this allows for a compact presentation of the results without any additional assumptions on the forward model except for continuity. Nevertheless, the analytic results of Section 4 also allow to show well-posedness for non-linear inverse problems under standard assumptions on the forward model such as in [32]. In Section 7, we develop and analyze a bilevel framework for learning the weights of regularization graphs. In particular, we show well-posedness and an example for a bilevel approach that allows to select optimal regularizers from a set of possible choices by learning zero-weights in the graph. An appendix further provides a list that shows how a selection of existing regularization functionals can be represented by regularization graphs.

Acknowledgements

2   Notation and assumptions

In this section we define the underlying setting and assumptions used in the paper. The structure of a general regularization functional will be represented by a directed graph $G = (V, E)$, where V is a non-empty finite set of nodes not containing 0 and $E \subset (V \times V) \setminus \{(n, n) : n \in V\}$ are the edges. We assume that G has a tree structure and that a root node $\hat{n} \in V$ exists, i.e., we assume that G contains no cycles and that for each $n \in V$ there exist edges $((n_{i-1}, n_i))_{i=1}^{M}$ in E such that $n_M = n$ and $n_0 = \hat{n}$.

We call a set $F \subset E$ a chain (of length $M > 0$ with root n_0) if $F = \{(n_{i-1}, n_i) \mid i = 1, \ldots, M, \ n_i \neq n_j \text{ for } i \neq j\}$. Further, for $n \in V$, we denote by n^- the node such that $(n^-, n) \in E$ if n is not the root node of the graph and $n^- = 0$ otherwise, noting that n^- is well defined due to the tree structure of G.

To any graph $G = (V, E)$ we associate a family of Banach spaces spaces $(X_n)_{n \in V}$ with the nodes and a family of Banach spaces $(X_e)_{e \in E}$ with the edges. Further, we associate the following functionals and operators with G.

- A convex functional $\Psi_n : X_n \to [0, \infty]$ for every $n \in V$.

- A linear *forward* operator $\Theta_{(n,m)} : \mathrm{dom}(\Theta_{(n,m)}) \subset X_{(n,m)} \to X_m$ for every $(n, m) \in E$.

- A linear *backward* operator $\Phi_{(n,m)} : X_{(n,m)} \to X_n$ for every $(n, m) \in E$.

We suppose that each X_n, $n \in V$ and each X_e, $e \in E$ admits a predual space denoted by $X_n^{\#}$ and $X_e^{\#}$, respectively, and make the following assumptions on $(\Psi_n)_n$, $(\Theta_e)_e$ and $(\Phi_e)_e$:

(H1) Ψ_n is weak* lower-semicontinuous for every $n \in V$.

(H2) For every $n \in V$, Ψ_n is coercive, i.e., for any sequence $(v^k)_k$ in X_n it holds

$$\|v^k\|_{X_n} \to +\infty \quad \Rightarrow \quad \Psi_n(v^k) \to +\infty \qquad \text{as} \qquad k \to +\infty.$$

(H3) $\Psi_n(0) = 0$ for every $n \in V$.

(H4) Θ_e is weak* closed for every $e \in E$.

(H5) $\ker(\Theta_e)$ is finite dimensional for every $e \in E$.

(H6) For every $e = (n, m) \in E$ there exists $C > 0$ and a continuous, linear projection $P_{\ker(\Theta_e)} : X_e \to \ker(\Theta_e)$ such that

$$\|w - P_{\ker(\Theta_e)}w\|_{X_e} \leq C\|\Theta_e w\|_{X_m} \tag{2.1}$$

for every $w \in \mathrm{dom}(\Theta_e)$.

(H7) Φ_e is weak* to weak* continuous for every $e \in E$.

(H8) Bounded sequences in X_e and X_n admit weak* convergent subsequences for every $e \in E$, $n \in V$.

Remark 2.1. We can observe the following details in the above assumptions:

- Hypothesis (H5) implies the existence of a linear and continuous projection on $\ker(\Theta_e)$.

- Under Hypothesis (H4), (H5) and (H8), Hypothesis (H6) is equivalent to Θ_e having closed range, see Lemma A.1 in the Appendix for a proof.

- Hypothesis (H7) implies the existence of a continuous predual operator for Φ_e for each $e \in E$. Consequently, each Φ_e is continuous as well (see for instance [10, Remark 3.2]).

- Hypothesis (H8) holds whenever X_e and X_n are reflexive or dual spaces of separable spaces. In case of reflexivity, the notion of weak* convergence can be replaced by weak convergence in all assumptions.

- Note that, since the Ψ_n are convex, assumption (H3) implies that $\Psi_n(\lambda v) \leq \lambda \Psi_n(v)$ for any $v \in X_n$, $\lambda \in (0, 1]$ and $n \in V$. This consequence of assumption (H3) will be needed in the context of varying the weights of a regularization graph. For well-posedness results such as existence and stability as presented in this paper, however, assumption (H3) is not necessary and could be dropped.

We also note that Hypothesis (H2) implies a coercivity estimate as follows.

Remark 2.2. Hypothesis (H2) holds if and only if there exists $C > 0$ and $D \in \mathbb{R}$ such that $\|v\|_{X_n} \leq C\Psi_n(v) + D$ for every $v \in X_n$. A proof for this can be found for example in [4, Fact 4.4.8].

We are now in a position to define the main objects of interest in this paper: Regularization graphs and associated regularization functionals. To this aim, we allow for weights of the form $(\alpha_e)_{e \in E}$ with $\alpha_e \in [0, \infty)$ for all $e \in E$.

Definition 2.3 (Regularization graph and associated regularization functional)**.** *Given* $G = (V, E)$ *a directed graph with tree structure and root node* $\hat{n}$*, and the associated spaces, functionals and operators as in Section 2 such that the hypotheses (H1) to (H8) hold, the* structure of a regularization graph *is defined as the tuple* $\mathcal{G} = (G, (\Psi_n)_{n \in V}, (\Theta_e)_{e \in E}, (\Phi_e)_{e \in E})$*. Together with a family of weights* $\alpha = (\alpha_e)_{e \in E}$*, a* regularization graph *is then defined as the tuple* $\mathcal{G}_\alpha = (\mathcal{G}, \alpha)$*.*

For any such regularization graph $\mathcal{G}_\alpha$, the associated regularization functional $R_\alpha = R(\mathcal{G}_\alpha) : X_{\hat{n}} \to [0, \infty]$ (called regularization graph functional) is defined as

$$R_\alpha(u) = \inf_{\substack{(w_e)_{e \in E} \\ w_e \in \mathrm{dom}(\Theta_e)}} \sum_{n \in V} \Psi_n\left(\Theta_{(n^-,n)} w_{(n^-,n)} - \sum_{(n,m) \in E} \alpha_{(n,m)} \Phi_{(n,m)} w_{(n,m)}\right)$$

$$= \inf\left\{ \sum_{n \in V} \Psi_n(v_n) \ \middle|\ \textit{for all } n \in V, \ e \in E \textit{ there exist } w_e \in \mathrm{dom}(\Theta_e) : \right.$$

$$\left. v_n = \Theta_{(n^-,n)} w_{(n^-,n)} - \sum_{(n,m) \in E} \alpha_{(n,m)} \Phi_{(n,m)} w_{(n,m)} \right\}$$

where we set $\Theta_{(\hat{n}^-,\hat{n})} = \mathrm{Id}$ and $w_{(\hat{n}^-,\hat{n})} = u$. Note that, in case of a trivial regularization graph, i.e., $V = \{\hat{n}\}$, $E = \emptyset$, we set $R_\alpha(u) = \Psi_{\hat{n}}(u)$.

Remark 2.4 (Weights). Generically, to each edge e within the graph structure of a regularization graph is associated a weight α_e. In many cases, e.g., when node functionals only take values in $\{0, \infty\}$, this leads to an overparametrization of the associated regularization functional. To avoid this, we often fix a subset of weights to be equal to 1 already when defining a regularization graph. Such weights are called *trivial weights*, and the other, non-trivial weights that might still vary are often numbered independently of the edge they are associated with.

Remark 2.5 (Graphical representation of regularization graphs). Let us revisit the prototypical graphical representation of a regularization graph in Figure 1. There, the circles represent nodes, with the node space shown above the circle and the functional Ψ_n inside. A splitting node is represented by a $\oplus$ and is associated with the functional $\mathcal{I}_{\{0\}}$. The rectangles denote the edges, with the edge space shown in the center, the forward operator Θ_e shown at the top and the backward operator Φ_e at the bottom. The weights $(\alpha_e)_{e \in E}$ are depicted as scalar factors in front of the backward operators $(\Phi_e)_{e \in E}$ (with arbitrary numbering independent of their position in the graph), and we use the convention that omitted weights at an edge e correspond to trivial weights $\alpha_e = 1$.
The arrows connect the nodes. At each node n, the node functional Ψ_n is evaluated at Θ_e of the variable from the incoming edge e minus the sum of all Φ_e applied to the variables w_e from the outgoing edges $\{e = (n, m) \in E : m \in V\}$. The regularization graph functional is given by minimizing this construction over all edge variables in the domain of the corresponding operators $(\Theta_e)_{e \in E}$.

We can also obtain a more compact representation of R_α as follows. Define the spaces

$$X_V = \bigtimes_{n \in V} X_n \quad \text{and} \quad X_E = \bigtimes_{e \in E} X_e$$

equipped with the product norm, and the operator $\Lambda_\alpha : \mathrm{dom}(\Lambda_\alpha) \subset X_E \to X_V$ as

$$(\Lambda_\alpha w)_n := \begin{cases} \Theta_{(n^-,n)} w_{(n^-,n)} - \sum_{(n,m) \in E} \alpha_{(n,m)} \Phi_{(n,m)} w_{(n,m)} & \text{for } n \in V \setminus \{\hat{n}\}, \\ -\sum_{(\hat{n},m) \in E} \alpha_{(\hat{n},m)} \Phi_{(\hat{n},m)} w_{(\hat{n},m)} & \text{for } n = \hat{n} \end{cases} \tag{2.2}$$

for every $n \in V$ and $w = (w_e)_{e \in E} \in \operatorname{dom}(\Lambda_\alpha)$ where $\operatorname{dom}(\Lambda_\alpha) = \bigtimes_{e \in E} \operatorname{dom}(\Theta_e)$. Then we can write the functional R_α associated with the regularization graph $\mathcal{G}_\alpha$ as

$$R_\alpha(u) = \inf \left\{ \Psi_{\hat{n}}(u + (\Lambda_\alpha w)_{\hat{n}}) + \sum_{n \in V \setminus \{\hat{n}\}} \Psi_n((\Lambda_\alpha w)_n) \,\middle|\, w \in \operatorname{dom}(\Lambda_\alpha) \right\}.$$

For notational convenience we also define the functional $\Psi_u : X_V \to [0, +\infty]$ for $u \in X_{\hat{n}}$ as

$$\Psi_u(v) := \Psi_{\hat{n}}(u + v_{\hat{n}}) + \sum_{n \in V \setminus \{\hat{n}\}} \Psi_n(v_n) \tag{2.3}$$

such that

$$R_\alpha(u) = \inf \left\{ \Psi_u(v) \,|\, v \in \operatorname{rg}(\Lambda_\alpha) \right\}. \tag{2.4}$$

Proposition 2.6. *Every regularization graph functional $R_\alpha : X_{\hat{n}} \to [0, +\infty]$ is convex, $R_\alpha(0) = 0$ and $R_\alpha(\lambda u) \leq \lambda R_\alpha(u)$ for all $u \in X_{\hat{n}}$, $\lambda \in (0, 1]$. Further, in case each Ψ_n for $n \in V$ is positively one homogeneoous, also R_α is positively one homogeneoous.*

The statement follows easily from the representation in (2.4) together with Assumption (H3).

2.1 Examples

In this section, we provide some concrete examples of regularization graphs to which our general assumptions apply. Here, for $d \in \mathbb{N}$, $d \geq 1$, we always denote by $\Omega \subset \mathbb{R}^d$ a bounded Lipschitz domain. Moreover, we denote by I the embedding of a Banach space into another one. We remark that domain and codomain of the embeddings change for different examples. However, they can easily be deduced from the context.
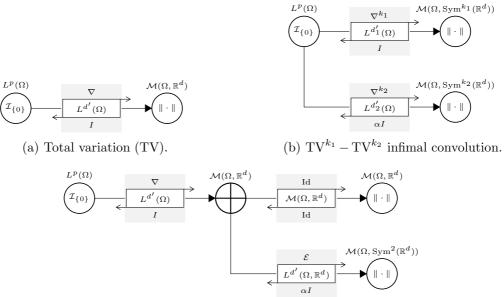
Total variation. Figure 2a shows the regularization graph corresponding to the total variation functional. The exponents for the Lebesgue spaces are chosen as $d' = d/(d-1)$ in case $d > 1$, $d' = \infty$ else, and $1 < p \leq d'$. Thanks to the embedding $\operatorname{BV}(\Omega) \hookrightarrow L^{d'}(\Omega)$, we set $\operatorname{BV}(\Omega)$ as the domain of the linear operator $\nabla : L^{d'}(\Omega) \to \mathcal{M}(\Omega, \mathbb{R}^d)$. With these choices, we now verify assumptions (H1)–(H8). Hypotheses (H1)–(H2) follow immediately from the weak* lower semicontinuity and coercivity of the total variation in $\mathcal{M}(\Omega, \mathbb{R}^d)$ [1]. It can easily be verified that the linear operator $\nabla : \operatorname{BV}(\Omega) \subset L^{d'}(\Omega) \to \mathcal{M}(\Omega, \mathbb{R}^d)$ is weak* closed and its kernel is the set of constant functions implying (H4) and (H5). Assumption (H6) follows from the Poincaré inequality for TV [1] and (H8) is a consequence of the Banach-Alaoglu theorem. Finally, the embedding $I : L^{d'}(\Omega) \to L^p(\Omega)$ is weak*-to-weak* continuous thanks to the exponents choice $1 < p \leq d'$, verifying assumption (H7).

The functional $R_\alpha : L^p(\Omega) \to [0, +\infty]$ associated to the regularization graph depicted in Figure 2a is given as
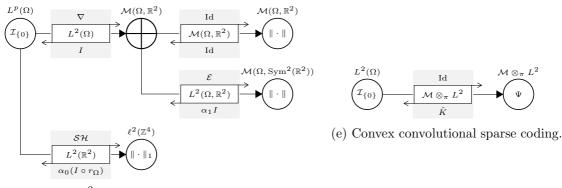
$$R_\alpha(u) = \inf_{w \in \operatorname{BV}(\Omega)} \mathcal{I}_{\{0\}}(u - w) + \|\nabla w\|_{\mathcal{M}} = \|\nabla u\|_{\mathcal{M}}$$

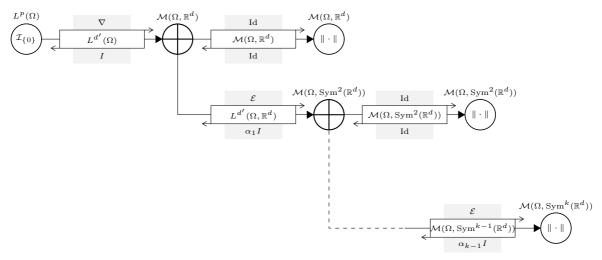for every $u \in \operatorname{BV}(\Omega)$ and $+\infty$ otherwise.

Infimal convolution of $\operatorname{TV}^{k_1} - \operatorname{TV}^{k_2}$. Figure 2b shows the regularization graph corresponding to the infimal convolution of TV^{k_1} and TV^{k_2} with $k_1, k_2 \in \mathbb{N}$. Here the exponents for the Lebesgue spaces are chosen as $d_i' = d/(d - k_i)$ in case $d_i > k_i$, $d_i' = \infty$ else, and $1 < p \leq \min\{d_1', d_2'\}$. Thanks to the embeddings $\operatorname{BV}^{k_i}(\Omega) \hookrightarrow L^{d_i'}(\Omega)$ we set $\operatorname{BV}^{k_i}(\Omega)$ as the

(a) Total variation (TV).

(b) $\mathrm{TV}^{k_1} - \mathrm{TV}^{k_2}$ infimal convolution.

(c) Second order total generalized variation.

(d) TGV^2-shearlet infimal convolution.

(e) Convex convolutional sparse coding.

(f) Total generalized variation of order k.

Figure 2: Examples of regularization graphs reproducing existing regularization functionals.

9

domains of the linear operators $\nabla^{k_i} : L^{d'_i}(\Omega) \to \mathcal{M}(\Omega, \mathrm{Sym}^{k_i}(\mathbb{R}^d))$, where $\mathrm{Sym}^k(\mathbb{R}^d)$ denotes the space of symmetric tensors of order k, e.g., $\mathbb{R}^d$ for $k = 1$ and the space of symmetric $d \times d$ matrices for $k = 2$. We refer to [7] for details and basic properties of $\mathrm{BV}^k(\Omega)$ and TV^k. By similar arguments to those used in the previous example and the generalized Poincaré inequality for TV^{k_i} [7, Corollary 3.23], it follows that our general assumptions (H1)–(H8) are satisfied. The functional $R_\alpha : L^p(\Omega) \to [0, +\infty]$ associated to the regularization graph depicted in Figure 2b is given as

$$R_\alpha(u) = \inf_{w_i \in \mathrm{BV}^{k_i}(\Omega)} \mathcal{I}_{\{0\}}(u - w_1 - \alpha w_2) + \|\nabla^{k_1} w_1\|_{\mathcal{M}} + \|\nabla^{k_2} w_2\|_{\mathcal{M}}$$
$$= \inf_{u = w_1 + \alpha w_2} \mathrm{TV}^{k_1}(w_1) + \mathrm{TV}^{k_2}(w_2).$$

Total generalized variation. Figure 2c shows the regularization graph corresponding to TGV^2_α, the second order TGV functional as in [8]. The exponents for the Lebesgue spaces are chosen as $d' = d/(d-1)$ in case $d > 1$, $d' = \infty$ else, and $1 < p \le d'$. The domain of the linear operator $\nabla : L^{d'}(\Omega) \to \mathcal{M}(\Omega, \mathbb{R}^d)$ is $\mathrm{BV}(\Omega)$ and the domain of the symmetrized gradient $\mathcal{E} : L^{d'}(\Omega, \mathbb{R}^d) \to \mathcal{M}(\Omega, \mathrm{Sym}^2(\mathbb{R}^d))$ is $\mathrm{BD}(\Omega)$, the space of functions of bounded deformation, where again we take advantage of the embeddings $\mathrm{BV}(\Omega) \hookrightarrow L^{d'}(\Omega)$ and $\mathrm{BD}(\Omega) \hookrightarrow L^{d'}(\Omega, \mathbb{R}^d)$. By similar arguments to those used the previous examples and the generalized Poincaré inequality for $w \mapsto \|\mathcal{E}w\|_{\mathcal{M}}$ [5, Corollary 4.20] it follows that in this setting our general assumptions (H1)–(H8) are satisfied. The functional $R_\alpha : L^p(\Omega) \to [0, +\infty]$ associated to the regularization graph functional depicted in Figure 2c is given as

$$R_\alpha(u) = \inf_{\substack{w \in \mathrm{BV}(\Omega), \\ w_1 \in \mathcal{M}(\Omega, \mathbb{R}^d), w_2 \in \mathrm{BD}(\Omega, \mathbb{R}^d)}} \mathcal{I}_{\{0\}}(u - w) + \mathcal{I}_{\{0\}}(\nabla w - w_1 - \alpha w_2) + \|w_1\|_{\mathcal{M}} + \|\mathcal{E}w_2\|_{\mathcal{M}}$$
$$= \inf_{w \in \mathrm{BD}(\Omega)} \|\nabla u - \alpha w\|_{\mathcal{M}} + \|\mathcal{E}w\|_{\mathcal{M}}$$

for $u \in \mathrm{BV}(\Omega)$ and $+\infty$ otherwise. Building on results in [6], also the TGV functional of arbitrary order $k \in \mathbb{N}$ can be realized via a regularization graph as in Figure 2f.

TGV^2-**shearlet infimal convolution.** Figure 2d shows the regularization graph that recovers a TGV^2-shearlet infimal convolution model introduced in [28] (see also [30]). Here $\Omega \subset \mathbb{R}^2$ is a bounded Lipschitz domain. The exponent for the Lebesgue space $L^p(\Omega)$ is chosen as $1 < p \le 2$. The domain of the linear operator $\nabla : L^2(\Omega) \to \mathcal{M}(\Omega, \mathbb{R}^2)$ is $\mathrm{BV}(\Omega)$ and the domain of the symmetrized gradient $\mathcal{E} : L^2(\Omega, \mathbb{R}^2) \to \mathcal{M}(\Omega, \mathrm{Sym}^2(\mathbb{R}^2))$ is $\mathrm{BD}(\Omega)$, where again we take advantage of the embeddings $\mathrm{BV}(\Omega) \hookrightarrow L^2(\Omega)$ and $\mathrm{BD}(\Omega) \hookrightarrow L^2(\Omega, \mathbb{R}^2)$. By similar arguments to those used the previous examples and the generalized Poincaré inequality for $w \mapsto \|\mathcal{E}w\|_{\mathcal{M}}$ [5, Corollary 4.20] it follows that in this setting our general assumptions (H1)–(H8) are satisfied for edges and nodes realizing the total generalized variation.

In order to introduce the shearlet transform in $L^2(\mathbb{R}^2)$ we start with several notations. First, for $a > 0$ and $s \in \mathbb{R}$ let A_a and S_s be the dilatation matrix and the shearing matrix defined respectively as

$$A_a = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}, \qquad S_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}.$$

The discrete shearlet system of $\Psi \in L^2(\mathbb{R}^2)$ is defined as

$$\Psi_{j,k,m}(x) = 2^{\frac{3}{4}j} \Psi(S_k A_{2^j}(x - m))$$

for $k, j \in \mathbb{Z}$ and $m \in \mathbb{Z}^2$ [36, Definition 8]. This allows to define the discrete shearlet transform operator $\mathcal{SH}$ as

$$\mathcal{SH}f(j, k, m) = \langle f, \Psi_{j,k,m} \rangle_{L^2} \tag{2.5}$$

for $f \in L^2(\mathbb{R}^2)$. By standard results in shearlet theory it holds that if Ψ is a classical shearlet, then $\mathcal{SH} : L^2(\mathbb{R}^2) \to \ell^2(\mathbb{Z}^4)$ is a Parseval frame for $L^2(\mathbb{R}^2)$, that is equivalent to $\|\mathcal{SH}f\|_{\ell^2(\mathbb{Z}^4)} = \|f\|_{L^2(\mathbb{R}^2)}$ for every $f \in L^2(\mathbb{R}^2)$ [36, Proposition 2]. In particular, this verifies (H6) for $\mathcal{SH}$. Moreover, a simple computation using that $\|\Psi_{j,k,m}\|_{L^2(\mathbb{R}^2)} = \|\Psi\|_{L^2(\mathbb{R}^2)}$ for every j, k, m together with Hölder's inequality shows that $\mathcal{SH}$ is weak*-to-weak* continuous, implying (H4). The backward operator $I \circ r_\Omega$ is the composition of the embedding $I : L^2(\Omega) \to L^p(\Omega)$ with the restriction $r_\Omega : L^2(\mathbb{R}^2) \to L^2(\Omega)$. It is immediate to check that $I \circ r_\Omega$ is weak*-to-weak* continuous showing (H7). Finally, we remark that the functional $\| \cdot \|_1 : \ell^2(\mathbb{Z}^4) \to [0, +\infty]$ is intended as the extension to $+\infty$ of the ℓ^1-norm on ℓ^2. Such extension is convex, coercive and weak* lower semicontinuous showing (H1)–(H3). The functional $R_\alpha : L^p(\Omega) \to [0, +\infty]$ associated to the regularization graph functional depicted in Figure 2d is then given as

$$\begin{aligned} R_\alpha(u) &= \inf_{\substack{w_1 \in \mathrm{BV}(\Omega), w_2 \in L^2(\mathbb{R}^2), \\ w_3 \in \mathcal{M}(\Omega, \mathbb{R}^2), w_4 \in \mathrm{BD}(\Omega)}} \mathcal{I}_{\{0\}}(u - w_1 - \alpha_0 r_\Omega w_2) + \mathcal{I}_{\{0\}}(\nabla w_1 - w_3 - \alpha_1 w_4) \\ &\qquad\qquad + \|w_3\|_{\mathcal{M}} + \|\mathcal{E}w_4\|_{\mathcal{M}} + \|\mathcal{SH}w_2\|_1 \\ &= \inf_{\substack{w_2 \in L^2(\mathbb{R}^2), \\ w_4 \in \mathrm{BD}(\Omega)}} \|\nabla(u - \alpha_0 r_\Omega w_2) - \alpha_1 w_4\|_{\mathcal{M}} + \|\mathcal{E}w_4\|_{\mathcal{M}} + \|\mathcal{SH}w_2\|_1. \end{aligned}$$

Convex convolutional sparse coding. Figure 2e shows the regularization graph corresponding to a data-adaptive convolutional-sparse-coding-based method recently introduced in [17]. As such methods are in general non-convex, in [17], the authors proposed a convex relaxation of the convolution LASSO problem in the tensor product of convolutional filter kernels and coefficient images. We refer to [17] for a more detailed description of the model. We denote by $\mathcal{M} \otimes_\pi L^2$ the projective tensor product between $\mathcal{M}(\Omega_\Sigma)$ and $L^2(\Sigma)$ [17, Appendix A], where Σ is a bounded Lipschitz domain and $\Omega_\Sigma := \Omega + \Sigma \subset \mathbb{R}^d$ is the Minkowski sum of Ω and Σ. The operator $\hat{K} : \mathcal{M} \otimes_\pi L^2 \to L^2(\Omega)$ is the unique tensor lifting of the bilinear operator $K : \mathcal{M}(\Omega_\Sigma) \times L^2(\Sigma) \to L^2(\Omega)$ defined essentially as

$$K(\mu, \theta)(x) = \int_{\Omega_\Sigma} \theta(x - y) \, d\mu(y). \tag{2.6}$$

Thanks to [17, Lemma 2], the operator $\hat{K}$ is weak* to weak* continuous. We also define the convex functional $\Psi : \mathcal{M} \otimes_\pi L^2 \to [0, +\infty]$ as $\Psi(C) = \|C\|_\pi + \nu\|C\|_{\mathrm{nuc}}$ for every $C \in \mathcal{M} \otimes_\pi L^2$, where $\nu > 0$ is a parameter,

$$\|C\|_\pi = \inf \left\{ \sum_{i=1}^\infty \|\mu_i\|_{\mathcal{M}} \|\theta_i\|_{L^2} \,\middle|\, C = \sum_{i=1}^\infty \mu_i \otimes_\pi \theta_i \right\} \tag{2.7}$$

is the projective norm of $\mathcal{M} \otimes_\pi L^2$ and

$$\|C\|_{\mathrm{nuc}} = \begin{cases} \sum_{i=1}^\infty \sigma_i(T_C) & \text{if } C \in L^2(\Omega_\Sigma) \otimes_\pi L^2(\Sigma), \\ +\infty & \text{otherwise,} \end{cases} \tag{2.8}$$

11

is an extension of the nuclear norm, where $\sigma_i(T_C)$ are the singular values of C interpreted as a bounded linear map from $L^2(\Omega_\Sigma)$ to $L^2(\Sigma)$. By Lemma 1 and Lemma 7 in [17] it follows that Ψ is weak* lower semicontinuous. Hence the general assumptions (H1)–(H8) for a regularization graph are satisfied and the functional $R_\alpha : L^2(\Omega) \to [0, +\infty]$ associated to the regularization graph depicted in Figure 2e is given as

$$R_\alpha(u) = \inf_{C \in \mathcal{M} \otimes_\pi L^2} \mathcal{I}_{\{0\}}(u - \hat{K}C) + \|C\|_\pi + \|C\|_{\mathrm{nuc}}.$$

Remark 2.7. The regularization graph functional for TV, infimal convolution of $\mathrm{TV}^{k_1} - \mathrm{TV}^{k_2}$ and total generalized variation can be extended to $L^1(\Omega)$ even if $L^1(\Omega)$ does not admit a predual. Such extension is described for general regularization graphs in Proposition 4.8.

3 Algebraic properties of regularization graphs

This section provides a recursive representation of regularization graphs and deals with estimates between different regularization graph functionals as well as their combination via addition or infimal convolution. First we need the definition of the height of a graph.

Definition 3.1 (Height of a regularization graph). *Given a regularization graph $\mathcal{G}_\alpha$ with $G = (V, E)$ the associated directed graph, we denote by $\mathcal{H}(\mathcal{G}_\alpha)$ its height defined as the number of edges in the longest path of G connecting the root to one of the leaves. That is, with $n_0 = \hat{n}$ the root node, we define*

$$\mathcal{H}(\mathcal{G}_\alpha) = \max\{M \,|\, \exists\, n_1, \ldots, n_M \text{ with } (n_{i-1}, n_i) \in E, \text{ for } i = 1, \ldots, M\}$$

if this set is non-empty and define $\mathcal{H}(\mathcal{G}_\alpha) = 0$ otherwise, i.e., in case of a trivial graph.

Note that the height of a regularization graph does not depend on the particular choice of weights α. Next, we provide a recursion result that allows us to rewrite a regularization graph of height h in terms of regularization graphs of height $h - 1$.

Lemma 3.2 (Recursive representation of regularization graphs). *For $\mathcal{G}_\alpha$ a regularization graph of height $h \geq 1$, $G = (V, E)$ the associated directed graph and $\hat{n}$ the root node, let $\hat{E} \subset E$ be the set all edges connected to the root node $\hat{n}$, $\hat{n}^{\hat{e}}$ for $\hat{e} \in \hat{E}$ be their endpoints and let $G^{\hat{e}} = (V^{\hat{e}}, E^{\hat{e}})$ be the subtree of $G = (V, E)$ with $\hat{n}^{\hat{e}}$ as root node.*
Then, there exist regularization graphs $\mathcal{G}^{\hat{e}}_{\alpha^{\hat{e}}}$ with associated directed graphs $G^{\hat{e}} = (V^{\hat{e}}, E^{\hat{e}})$ of height at most $h - 1$ and weights $(\alpha^{\hat{e}}_e)_{e \in E^{\hat{e}}}$ such that, with $R_\alpha = R(\mathcal{G}_\alpha)$ and $R^{\hat{e}}_{\alpha^{\hat{e}}} = R(\mathcal{G}^{\hat{e}}_{\alpha^{\hat{e}}})$ the associated functionals, the following recursive representation holds

$$R_\alpha(u) = \inf \left\{ \Psi_{\hat{n}}\left(u - \sum_{\hat{e} \in \hat{E}} \alpha_{\hat{e}} \Phi_{\hat{e}} w_{\hat{e}}\right) + \sum_{\hat{e} \in \hat{E}} R^{\hat{e}}_{\alpha^{\hat{e}}}(\Theta_{\hat{e}} w_{\hat{e}}) \,\Big|\, w_{\hat{e}} \in \mathrm{dom}(\Theta_{\hat{e}}) \text{ for all } \hat{e} \in \hat{E} \right\}. \quad (3.1)$$

Proof. We explicitly construct the claimed recursive representation as visualized in Figure 3. First note that we can re-write R_α as

$$R_\alpha(u) = \inf \left\{ \Psi_{\hat{n}}\left(u - \sum_{\hat{e} \in \hat{E}} \alpha_{\hat{e}} \Phi_{\hat{e}} w_{\hat{e}}\right) + \sum_{\hat{e} \in \hat{E}} \sum_{n \in V^{\hat{e}}} \Psi_n(v_n) \,\Big|\, w_e \in \mathrm{dom}(\Theta_e) \text{ for all } e \in E, \right.$$

$$\left. \forall \hat{e} \in \hat{E}, n \in V^{\hat{e}} : v_n = \Theta_{(n^-, n)} w_{(n^-, n)} - \sum_{(n, m) \in E^{\hat{e}}} \alpha_{(n, m)} \Phi_{(n, m)} w_{(n, m)} \right\}.$$
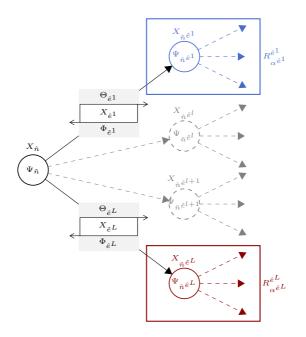
12

Figure 3: The recursive representation of a regularization graph according to Lemma 3.2.

Now define $\mathcal{G}^{\hat{e}}_{\alpha^{\hat{e}}}$ to be a regularization graph with graph structure $G^{\hat{e}} = (V^{\hat{e}}, E^{\hat{e}})$ and the associated operators, functionals and weights, such that $R^{\hat{e}}_{\alpha^{\hat{e}}} = R(\mathcal{G}^{\hat{e}}_{\alpha^{\hat{e}}})$ is given as

$$R^{\hat{e}}_{\alpha^{\hat{e}}}(z) = \inf \left\{ \sum_{n \in V^{\hat{e}}} \Psi_n(v_n) \,\middle|\, w_e \in \mathrm{dom}(\Theta_e) \text{ for all } e \in E^{\hat{e}}, \right.$$

$$\left. \forall n \in V^{\hat{e}} : v_n = \Theta_{(n^-,n)} w_{(n^-,n)} - \sum_{(n,m) \in E^{\hat{e}}} \alpha_{(n,m)} \Phi_{(n,m)} w_{(n,m)} \right\},$$

where we note that here $\hat{n}^{\hat{e}}$ is regarded as a node of $\mathcal{G}^{\hat{e}}_{\alpha^{\hat{e}}}$, and thus $\Theta_{((\hat{n}^{\hat{e}})^-,\hat{n}^{\hat{e}})} = \mathrm{Id}$ and $w_{((\hat{n}^{\hat{e}})^-,\hat{n}^{\hat{e}})} = z$. The recursive representation of R_α is then given as

$$R_\alpha(u) = \inf \left\{ \Psi_{\hat{n}}\left(u - \sum_{\hat{e} \in \hat{E}} \alpha_{\hat{e}} \Phi_{\hat{e}} w_{\hat{e}} \right) + \sum_{\hat{e} \in \hat{E}} \sum_{n \in V^{\hat{e}}} \Psi_n(v_n) \,\middle|\, w_e \in \mathrm{dom}(\Theta_e) \text{ for all } e \in E, \right.$$

$$\forall \hat{e} \in \hat{E}, \forall n \in V^{\hat{e}} \setminus \{\hat{n}^{\hat{e}}\} : v_n = \Theta_{(n^-,n)} w_{(n^-,n)} - \sum_{(n,m) \in E^{\hat{e}}} \alpha_{(n,m)} \Phi_{(n,m)} w_{(n,m)},$$

$$\left. v_{\hat{n}^{\hat{e}}} = \Theta_{\hat{e}} w_{\hat{e}} - \sum_{(\hat{n}^{\hat{e}},m) \in E^{\hat{e}}} \alpha_{(\hat{n}^{\hat{e}},m)} \Phi_{(\hat{n}^{\hat{e}},m)} w_{(\hat{n}^{\hat{e}},m)} \right\}$$

$$= \inf \left\{ \Psi_{\hat{n}}\left(u - \sum_{\hat{e} \in \hat{E}} \alpha_{\hat{e}} \Phi_{\hat{e}} w_{\hat{e}} \right) + \sum_{\hat{e} \in \hat{E}} R^{\hat{e}}_{\alpha^{\hat{e}}}(\Theta_{\hat{e}} w_{\hat{e}}) \,\middle|\, w_{\hat{e}} \in \mathrm{dom}(\Theta_{\hat{e}}) \text{ for all } \hat{e} \in \hat{E} \right\},$$

which proves the assertion. □

As first consequence of this recursive representation, we obtain an estimate between two regularization graph functionals corresponding to regularization graphs with different weights.

13

Lemma 3.3. *Let $\mathcal{G}_{\alpha^1}$ and $\mathcal{G}_{\alpha^2}$ be two regularization graphs with the same underlying graph structure $\mathcal{G}$ and directed graph $G = (V, E)$ with root node $\hat{n}$, and let α^1, α^2 be weights such that $\alpha_e^1 \geq \alpha_e^2$ for all $e \in E$. Then, with*

$$C_{\alpha^1,\alpha^2} := \max \Big\{ \prod_{e \in F} \frac{\alpha_e^2}{\alpha_e^1} \,\Big|\, F \subset E \text{ is either empty or a chain with } \hat{n} \text{ as root}\Big\}, \qquad (3.2)$$

where we use the conventions $\prod_{e \in \emptyset} \frac{\alpha_e^2}{\alpha_e^1} = 1$ and $\frac{0}{0} = 0$, for the associated regularization graph functionals $R_{\alpha^1} = R(\mathcal{G}_{\alpha^1})$ and $R_{\alpha^2} = R(\mathcal{G}_{\alpha^2})$ it holds that

$$R_{\alpha^1}(u) \leq C_{\alpha^1,\alpha^2} R_{\alpha^2}(u) \quad \text{for all } u \in X_{\hat{n}}.$$

Proof. We prove the result by induction over the height h of the graphs. Assume the result holds true for any two regularization graphs with height less than h.

Now note that, by assumption, $(\alpha_e^1 = 0)$ implies $(\alpha_e^2 = 0)$, so we can adapt the graph $G = (V, E)$ by removing all edges $e \in E$ with $\alpha_e^1 = 0$ and all subsequently disconnected nodes, without increasing its height, changing C_{α^1,α^2} or the values of the R_{α^i}. Hence, without loss of generality, assume that $\alpha_e^1 > 0$ for all $e \in E$. Now for $h = 0$ the result holds trivially and for $h \geq 1$ we can use the recursive representation of Lemma 3.2 to obtain

$$R_{\alpha^1}(u) = \inf\left\{ \Psi_{\hat{n}}\left(u - \sum_{\hat{e} \in \hat{E}} \Phi_{\hat{e}} w_{\hat{e}} \right) + \sum_{\hat{e} \in \hat{E}} R_{(\alpha^1)^{\hat{e}}}^{\hat{e}}\left(\frac{\Theta_{\hat{e}} w_{\hat{e}}}{\alpha_{\hat{e}}^1} \right) \,\Bigg|\, w_{\hat{e}} \in \mathrm{dom}(\Theta_{\hat{e}}) \text{ for all } \hat{e} \in \hat{E} \right\}$$

$$\leq \inf\left\{ \Psi_{\hat{n}}\left(u - \sum_{\hat{e} \in \hat{E}} \alpha_{\hat{e}}^2 \Phi_{\hat{e}} w_{\hat{e}} \right) + \sum_{\hat{e} \in \hat{E}} \frac{\alpha_{\hat{e}}^2}{\alpha_{\hat{e}}^1} R_{(\alpha^1)^{\hat{e}}}^{\hat{e}}(\Theta_{\hat{e}} w_{\hat{e}}) \,\Bigg|\, w_{\hat{e}} \in \mathrm{dom}(\Theta_{\hat{e}}) \text{ for all } \hat{e} \in \hat{E} \right\}$$

$$\leq C_{\alpha^1,\alpha^2} R_{\alpha^2}(u),$$

where in the first line we substituted $w_{\hat{e}}$ to $\alpha_{\hat{e}}^1 w_{\hat{e}}$ and in the second line we substituted $\alpha_{\hat{e}}^2 w_{\hat{e}}$ to $w_{\hat{e}}$; additionally, in the first inequality we used that $R_{(\alpha^1)^{\hat{e}}}^{\hat{e}}(\lambda u) \leq \lambda R_{(\alpha^1)^{\hat{e}}}^{\hat{e}}(u)$ for $\lambda \in [0,1]$, see Proposition 2.6, and we obtained the last estimate from the induction hypothesis and the definition of C_{α^1,α^2}. $\square$

Remark 3.4. It is easy to see from the proof above that, whenever Ψ_n for some $n \in V$ is positive one-homogeneous, the assumption $\alpha_e^1 \geq \alpha_e^2$ for $e = (n^-, n)$ can be replaced by $(\alpha_e^1 = 0)$ implying $(\alpha_e^2 = 0)$. In particular, if $(\alpha_e^1 = 0)$ if and only if $(\alpha_e^2 = 0)$ for all $e = (n^-, n)$ such that Ψ_n is positive one-homogeneous and all other α_e^1 and α_e^2 coincide, then R_{α^1} and R_{α^2} are equivalent, i.e., R_{α^1} can be estimated from above and below by a constant times R_{α^2}, $\mathrm{dom}(R_{\alpha^1}) = \mathrm{dom}(R_{\alpha^2})$ and also their zero-sets coincide.

3.1 Combining regularization graphs

Obviously, for $R_\alpha = R(\mathcal{G}_\alpha)$ being a regularization graph funtional and $\lambda > 0$, also λR_α is a regularization graph functional (corresponding to an adaption of the regularization graph $\mathcal{G}_\alpha$ where all node functionals Ψ_n are replaced by $\lambda \Psi_n$). In this subsection we show that also the sum and infimal-convolution of two regularization graph functionals are again regularization graph functionals.
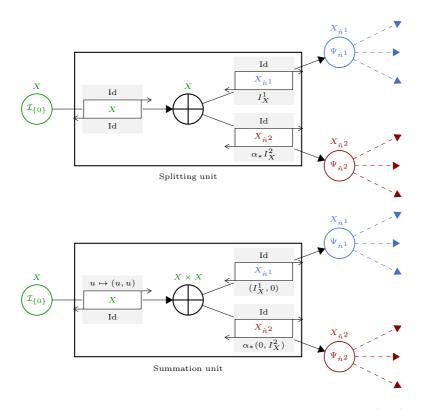
Figure 4: Combining regularization graphs via infimal convolution (top) and summation (bottom).

Proposition 3.5 (Infimal-convolution and sum of regularization graphs). *Let $\mathcal{G}^1_{\alpha^1} = (\mathcal{G}^1, \alpha^1)$ and $\mathcal{G}^2_{\alpha^2} = (\mathcal{G}^2, \alpha^2)$ be two regularization graphs with associated directed graphs $G^1 = (V^1, E^1)$ and $G^2 = (V^2, E^2)$ and root nodes $\hat{n}^1$ and $\hat{n}^2$, respectively. Let X be a Banach space admitting a predual such that bounded sequences in X admit weak* convergent subsequences and such that the embeddings $I^1_X : X_{\hat{n}^1} \to X$ and $I^2_X : X_{\hat{n}^2} \to X$ are continuous w.r.t. weak* convergence.*
For additional nodes $\hat{n}, \hat{n}^0 \notin V^1 \cup V^2$ define the combined graph $G = (V, E)$ with

$$V = \{\hat{n}, \hat{n}^0\} \cup V^1 \cup V^2, \quad E = \{(\hat{n}, \hat{n}^0), (\hat{n}^0, \hat{n}^1), (\hat{n}^0, \hat{n}^2)\} \cup E^1 \cup E^2.$$

Then, combined regularization graphs can be constructed as follows.

- ***Infimal-convolution:*** *For the additional nodes and edges in G, define the spaces*

$$X_{\hat{n}} = X_{\hat{n}^0} = X, \quad X_{(\hat{n}, \hat{n}^0)} = X, \quad X_{(\hat{n}^0, \hat{n}^l)} = X_{\hat{n}^l}, \quad l = 1, 2,$$

the operators

$$\Theta_{(\hat{n}, \hat{n}^0)} = \Theta_{(\hat{n}^0, \hat{n}^1)} = \Theta_{(\hat{n}^0, \hat{n}^2)} = \mathrm{Id}, \quad \Phi_{(\hat{n}, \hat{n}^0)} = \mathrm{Id}, \quad \Phi_{(\hat{n}^0, \hat{n}^1)} = I^1_X, \quad \Phi_{(\hat{n}^0, \hat{n}^2)} = I^2_X,$$

and the functionals and weights

$$\Psi_{\hat{n}} = \Psi_{\hat{n}^0} = \mathcal{I}_{\{0\}}, \quad \alpha_{(\hat{n}, \hat{n}^0)} = \alpha_{(\hat{n}^0, \hat{n}^1)} = 1, \quad \alpha_{(\hat{n}^0, \hat{n}^2)} = \alpha_* \in [0, \infty),$$

15

and adopt the elements of $\mathcal{G}^1_{\alpha^1}$ and $\mathcal{G}^2_{\alpha^2}$ for all other nodes and edges. Then, the associated structure $\mathcal{G}_\alpha = (\mathcal{G}, \alpha)$ defines a regularization graph and, for $R_\alpha = R(\mathcal{G}_\alpha)$, $R^1_{\alpha^1} = R(\mathcal{G}^1_{\alpha^1})$ and $R^2_{\alpha^2} = R(\mathcal{G}^2_{\alpha^2})$, it holds that

$$R_\alpha(u) = \inf_{v \in X_{\hat{n}}} R^1_{\alpha^1}(u - \alpha_* v) + R^2_{\alpha^2}(v).$$

- **Summation:** *For the additional nodes and edges in G, define the spaces*

$$X_{\hat{n}} = X, \quad X_{\hat{n}^0} = X \times X, \quad X_{(\hat{n}, \hat{n}^0)} = X, \quad X_{(\hat{n}^0, \hat{n}^l)} = X_{\hat{n}^l}, \quad l = 1, 2,$$

the operators

$$\Theta_{(\hat{n}, \hat{n}^0)} = [u \mapsto (u, u)], \quad \Theta_{(\hat{n}^0, \hat{n}^1)} = \Theta_{(\hat{n}^0, \hat{n}^2)} = \mathrm{Id},$$
$$\Phi_{(\hat{n}, \hat{n}^0)} = \mathrm{Id}, \quad \Phi_{(\hat{n}^0, \hat{n}^1)} = [u \mapsto (I^1_X u, 0)], \quad \Phi_{(\hat{n}^0, \hat{n}^2)} = [u \mapsto (0, I^2_X u)],$$

and the functionals and weights

$$\Psi_{\hat{n}} = \Psi_{\hat{n}^0} = \mathcal{I}_{\{0\}}, \quad \alpha_{(\hat{n}, \hat{n}^0)} = \alpha_{(\hat{n}^0, \hat{n}^1)} = 1, \quad \alpha_{(\hat{n}^0, \hat{n}^2)} = \alpha_* \in (0, \infty),$$

and adopt the elements of $\mathcal{G}^1_{\alpha^1}$ and $\mathcal{G}^2_{\alpha^2}$ for all other nodes and edges. Then, the associated structure $\mathcal{G}_\alpha = (\mathcal{G}, \alpha)$ defines a regularization graph and, for $R_\alpha = R(\mathcal{G}_\alpha)$, $R^1_{\alpha^1} = R(\mathcal{G}^1_{\alpha^1})$ and $R^2_{\alpha^2} = R(\mathcal{G}^2_{\alpha^2})$, it holds that

$$R_\alpha(u) = R^1_{\alpha^1}(u) + R^2_{\alpha^2}(\alpha_*^{-1} u).$$

Proof. It is easy to see that all spaces, functionals, operators and weights involved in the definition of $\mathcal{G}_\alpha$ fulfill Assumptions (H1) to (H8), such that $\mathcal{G}_\alpha$ defines a regularization graph. Denote the edges $e^l = (\hat{n}, \hat{n}^l)$ for $l \in \{0, 1, 2\}$. For the claimed representation of R_α in case of the infimal-convolution, we observe that

$$R_\alpha(u) = \inf \left\{ \mathcal{I}_{\{0\}}(u - w_{e^0}) + \mathcal{I}_{\{0\}}(w_{e^0} - w_{e^1} - \alpha_* w_{e^2}) + \sum_{l=1}^{2} \sum_{n^l \in V^l} \Psi_{n^l}(v_{n^l}) \,\Big|\right.$$

$$\forall l \in \{1, 2\}, n^l \in V^l \setminus \{\hat{n}^l\} : v_{n^l} = \Theta_{((n^l)^-, n^l)} w_{((n^l)^-, n^l)} - \sum_{(n^l, m) \in E^l} \alpha_{(n^l, m)} \Phi_{(n^l, m)} w_{(n^l, m)},$$

$$\left. v_{\hat{n}^l} = w_{e^l} - \sum_{(\hat{n}^l, m) \in E^l} \alpha_{(\hat{n}^l, m)} \Phi_{(\hat{n}^l, m)} w_{(\hat{n}^l, m)}, \forall e \in E : w_e \in \mathrm{dom}(\Theta_e) \right\}$$

$$= \inf \left\{ \sum_{l=1}^{2} \sum_{n^l \in V^l} \Psi_{n^l}(v_{n^l}) \,\Big|\, w_{e^1} + \alpha_* w_{e^2} = u, \forall l \in \{1, 2\}, n^l \in V^l \setminus \{\hat{n}^l\} : \right.$$

$$v_{n^l} = \Theta_{((n^l)^-, n^l)} w_{((n^l)^-, n^l)} - \sum_{(n^l, m) \in E^l} \alpha_{(n^l, m)} \Phi_{(n^l, m)} w_{(n^l, m)},$$

$$\left. v_{\hat{n}^l} = w_{e^l} - \sum_{(\hat{n}^l, m) \in E^l} \alpha_{(\hat{n}^l, m)} \Phi_{(\hat{n}^l, m)} w_{(\hat{n}^l, m)}, \forall e \in E : w_e \in \mathrm{dom}(\Theta_e) \right\}$$

$$= \inf_{v \in X_{\hat{n}}} R^1_{\alpha^1}(u - \alpha_* v) + R^2_{\alpha^2}(v).$$

16

Likewise, for the claimed representation of R_α in case of summation, we observe that

$$R_\alpha(u) = \inf \left\{ \mathcal{I}_{\{0\}}(u - w_{e^0}) + \mathcal{I}_{\{(0,0)\}}((w_{e^0} - w_{e^1}, w_{e^0} - \alpha_* w_{e^2})) + \sum_{l=1}^{2} \sum_{n^l \in V^l} \Psi_{n^l}(v_{n^l}) \right|$$

$$\forall l \in \{1,2\}, n^l \in V^l \setminus \{\hat{n}^l\} : v_{n^l} = \Theta_{((n^l)^-, n^l)} w_{((n^l)^-, n^l)} - \sum_{(n^l, m) \in E^l} \alpha_{(n^l, m)} \Phi_{(n^l, m)} w_{(n^l, m)},$$

$$v_{\hat{n}^l} = w_{e^l} - \sum_{(\hat{n}^l, m) \in E^l} \alpha_{(\hat{n}^l, m)} \Phi_{(\hat{n}^l, m)} w_{(\hat{n}^l, m)}, \forall e \in E : w_e \in \mathrm{dom}(\Theta_e) \right\}$$

$$= \inf \left\{ \sum_{l=1}^{2} \sum_{n^l \in V^l} \Psi_{n^l}(v_{n^l}) \right| w_{e^1} = u, w_{e^2} = \alpha_*^{-1} u, \forall l \in \{1,2\}, n^l \in V^l \setminus \{\hat{n}^l\} :$$

$$v_{n^l} = \Theta_{((n^l)^-, n^l)} w_{((n^l)^-, n^l)} - \sum_{(n^l, m) \in E^l} \alpha_{(n^l, m)} \Phi_{(n^l, m)} w_{(n^l, m)},$$

$$v_{\hat{n}^l} = w_{e^l} - \sum_{(\hat{n}^l, m) \in E^l} \alpha_{(\hat{n}^l, m)} \Phi_{(\hat{n}^l, m)} w_{(\hat{n}^l, m)}, \forall e \in E : w_e \in \mathrm{dom}(\Theta_e) \right\}$$

$$= R_{\alpha^1}^1(u) + R_{\alpha^2}^2(\alpha_*^{-1} u).$$

$\square$

More generally, note that any regularization graph can be extended by appending another regularization graph to one of its leaves, and in particular by appending a regularization graph corresponding to the infimal convolution or the sum of two other regularization graph functionals. The latter can be achieved by appending a splitting or summation unit as in Figure 4 to a leaf-node, where the $\mathcal{I}_{\{0\}}$ and X in the left, green nodes in Figure 4 are replaced by the corresponding node functional and node space of the leaf node.

Remark 3.6 (Assumptions on the sum of two regularization graphs). The notion of regularization graphs was designed mainly for an infimal-convolution-type combination of functionals and operators, since we believe this situation is more interesting in practice. For infimal-convolution-type combinations, we believe that our assumptions on the underlying functionals and operators are rather minimal. Our framework also allows for the summation of two functionals, but in this situation our assumptions are suboptimal. Indeed, one would expect that in a summation, only one of the two functionals needs to fulfill the assumptions of a regularization graph in order to provide a suitable regularization strategy. Indeed, when using the sum of two (suitable) functionals for regularization, generically, only one of them needs to fulfill coercivity properties (such as (H2) together with (H6)) in order to obtain well-posedness results for linear inverse problems. Nevertheless, we do not further generalize our framework towards weakening the assumptions for the sum of two functionals since i) we believe this situation is less relevant and ii) this would significantly complicate our basic assumptions and results for instance on convergence for vanishing noise and bilevel optimization, thereby hindering our main goal of providing an easily applicable framework.

4 Analytic properties of regularization graphs

The goal of this section is to obtain analytic properties of regularization graph functionals that provide the basis for well-posedness results for the regularization of inverse problems. To this aim, we first consider lower semi-continuity and coercivity properties, for which we need a general lemma that deals with projections.

For the lemma, remember that for a Banach space X and a finite dimensional subspace L of X, there always exists a bounded linear projection $P : X \to L$.

Lemma 4.1. *Let X be a Banach space, $R : X \to [0, \infty]$ be a functional, $L \subset X$ be a finite dimensional subspace and assume there is a function $G : X \to L$ and $C > 0, D \geq 0$ such that*

$$\|u - G(u)\|_X \leq CR(u) + D$$

for all $u \in X$. Then, for a closed subspace $K \subset X$ and a bounded, linear projection $P_{K \cap L} : X \to K \cap L$ there exist constants $\tilde{C} > 0, \tilde{D} \geq 0$, with $\tilde{D} = 0$ in case $D = 0$, such that for all $u \in K$,

$$\|u - P_{K \cap L}u\|_X \leq \tilde{C}R(u) + \tilde{D}. \tag{4.1}$$

In particular, if $K = X$, this holds for any bounded linear projection onto L.

Proof. Assume this does not hold true, then we can pick a sequence $(u^k)_k$ in K such that

$$\|u^k - P_{K \cap L}u^k\|_X > kR(u^k) + kD/C$$

for each k, with C, D being the constants of the original estimate. This implies in particular that $\|u^k - P_{K \cap L}u^k\|_X > 0$ for each k. Defining $\tilde{u}^k = (u^k - P_{K \cap L}u^k)/\|u^k - P_{K \cap L}u^k\|_X$ such that $\|\tilde{u}^k\|_X = 1$, we can estimate

$$C/k \geq C\frac{R(u^k) + D/C}{\|u^k - P_{K \cap L}u^k\|_X} \geq \frac{\|u^k - G(u^k)\|_X}{\|u^k - P_{K \cap L}u^k\|_X} = \|\tilde{u}^k - \frac{G(u^k) - P_{K \cap L}u^k}{\|u^k - P_{K \cap L}u^k\|_X}\|_X.$$

In particular, this implies that $((G(u^k) - P_{K \cap L}u^k)/\|u^k - P_{K \cap L}u^k\|_X)_k$ is bounded and, by finite dimensionality, admits a (non-relabeled) subsequence strongly converging to some $z \in L$. Consequently, also $(\tilde{u}^k)_k$ converges strongly to z and from closedness of K we get that $z \in K$. Hence

$$0 = (\mathrm{Id} - P_{K \cap L})z = \lim_k (\mathrm{Id} - P_{K \cap L})\tilde{u}^k = \lim_k \tilde{u}^k$$

strongly, which is a contradiction to $\|\tilde{u}^k\|_X = 1$ for each k and concludes the proof of the coercivity estimate in the general form. Also, it can be seen from this argument that we can choose $\tilde{D} = 0$ in case $D = 0$, which completes the proof. $\qquad\square$

The following lemma provides a standard lower semi-continuity and compactness result. Since it will be used frequently in the paper, we provide its proof for the sake of completeness.

Lemma 4.2. *Let X, W and Y be Banach spaces such that bounded sequences in X and W admit weak*-convergent subsequences. For $\Theta : \mathrm{dom}(\Theta) \subset X \to W$ a weak* to weak* closed operator and $F : W \to [0, \infty)$ convex and weak* lower semi-continuous, suppose that:*

i) There exists a finite dimensional subspace $L \subset W$ such that $F(u + v) = F(u)$ for all $u \in W$, $v \in L$, a continuous, linear projection $P_L : W \to L$ and $C > 0, D \geq 0$ such that

$$\|z - P_L z\|_X \leq CF(z) + D \qquad \forall z \in W. \tag{4.2}$$

18

ii) The space $\ker(\Theta)$ *is finite dimensional and there exists a continuous projection* $P_{\ker(\Theta)}$: $X \to \ker(\Theta)$ *and* $B > 0$ *such that*

$$\|u - P_{\ker(\Theta)}u\|_X \leq B\|\Theta u\|_W \qquad \forall u \in \operatorname{dom}(\Theta) \subset X. \tag{4.3}$$

Then, defining $F \circ \Theta : X \to [0, \infty]$ *as*

$$(F \circ \Theta)(u) = \begin{cases} F(\Theta u) & \textit{if } u \in \operatorname{dom}(\Theta), \\ \infty & \textit{else,} \end{cases}$$

implies that $F \circ \Theta$ *is weak* lower semi-continuous.*
Further, for $K : X \to Y$ *a linear, bounded operator, for any sequence* $(w^k)_k$ *in* X *such that both* $(Kw^k)_k$ *and* $((F \circ \Theta)(w^k))_k$ *are bounded, there exists a sequence* $(\tilde{w}^k)_k$ *such that* $w^k - \tilde{w}^k \in \ker(K) \cap \Theta^{-1}(L)$ *and both* $(\tilde{w}^k)_k$ *and* $(\Theta\tilde{w}^k)_k$ *are bounded.*

Proof. At first note that, by Lemma 4.1 and using that $\operatorname{rg}(\Theta)$ is closed (Lemma A.1) there exist $\tilde{C} > 0$, $\tilde{D} \geq 0$ such that

$$\|z - P_{\operatorname{rg}(\Theta) \cap L} z\|_W \leq \tilde{C}F(z) + \tilde{D} \quad \text{for all } z \in \operatorname{rg}(\Theta),$$

where $P_{\operatorname{rg}(\Theta) \cap L}$ is a bounded linear projection onto the finite dimensional space $\operatorname{rg}(\Theta) \cap L$. Defining $\ker(\Theta)^{\perp} := \operatorname{rg}(\operatorname{Id} - P_{\ker(\Theta)}) \cap \operatorname{dom}(\Theta)$, it is easy to see that $\ker(\Theta)^{\perp}$ is a complement of $\ker(\Theta)$ in $\operatorname{dom}(\Theta)$ and that Θ is injective on $\ker(\Theta)^{\perp}$. Hence, with $M := \Theta^{-1}(L)$, we can define $P : \operatorname{dom}(\Theta) \to M$ as

$$Pw := \Theta^{-1}P_{\operatorname{rg}(\Theta) \cap L}\Theta w + P_{\ker(\Theta)}w,$$

where Θ^{-1} is the inverse of $\Theta : \ker(\Theta)^{\perp} \to \operatorname{rg}(\Theta)$. Then, we observe that $F \circ \Theta$ is invariant on M and that M is a finite dimensional vector space. Further, for any $w \in \operatorname{dom}(\Theta)$,

$$\begin{aligned}
\|w - Pw\|_X &= \|w - \Theta^{-1}P_{\operatorname{rg}(\Theta) \cap L}\Theta w - P_{\ker(\Theta)}(w - \Theta^{-1}P_{\operatorname{rg}(\Theta) \cap L}\Theta w)\|_X \\
&\leq B\|\Theta(w - \Theta^{-1}P_{\operatorname{rg}(\Theta) \cap L}\Theta w)\|_W \\
&= B\|\Theta(w - Pw)\|_W \\
&= B\|\Theta w - P_{\operatorname{rg}(\Theta) \cap L}\Theta w\|_W \leq B\tilde{C}F(\Theta w) + B\tilde{D}.
\end{aligned} \tag{4.4}$$

We now prove that $F \circ \Theta$ is weak* lower semi-continuous. Take $(w^k)_k$ weak* converging to some $w \in X$. Without loss of generality, we can assume that $\liminf_k (F \circ \Theta)(w^k) < \infty$ and, up to extracting a subsequence, we can choose w^k such that it realizes the lim inf. Then, by the estimate (4.4) above, both $(\|w^k - Pw^k\|_X)_n$ and $(\|\Theta(w^k - Pw^k)\|_W)_k$ are bounded such that, by taking a non-relabeled subsequence, we can assume that $\Theta(w^k - Pw^k)$ weak* converges to some $z \in W$ and $w^k - Pw^k$ weak* converges to some $v \in X$. Weak* closedness of Θ then implies that $v \in \operatorname{dom}(\Theta)$ and $\Theta v = z$. Also, thanks to the finite dimensionality of M we have that $w - v = \text{w*-}\lim_k w^k - (w^k - Pw^k) \in M$ such that, by weak* lower semi-continuity of F, we conclude

$$F(\Theta w) = F(\Theta v) \leq \liminf_k F(\Theta(w^k - Pw^k)) = \liminf_k F(\Theta w^k)$$

as claimed.

Now take $(w^k)_k$ in X such that both $(Kw^k)_k$ and $(F \circ \Theta(w^k))_k$ are bounded. Let Z be a complement of $\ker(K) \cap M$ in M and $P_Z : M \to Z$ a continuous linear projection onto Z, i.e., $\mathrm{Id} - P_Z : M \to M$ is a projection onto $\ker(K) \cap M$. Then, define $\tilde{w}^k = w^k - (Pw^k - P_Z Pw^k)$. It follows that $w^k - \tilde{w}^k = (\mathrm{Id} - P_Z)Pw^k \in \ker(K) \cap M$ and that $(w^k - Pw^k)_k$ is bounded in X by the estimate (4.4) and the boundedness of $(F \circ \Theta(w^k))_k$. Now since K is injective on the finite dimensional space Z and $KP_Z = K$, we further get for $A > 0$ a generic constant that

$$\|P_Z Pw^k\|_X \leq A\|KP_Z Pw^k\|_Y \leq A(\|Kw^k\|_Y + \|K(w^k - Pw^k)\|_Y) < A < \infty$$

such that $(\tilde{w}^k)_k$ is bounded. Boundedness of Θ on the finite dimensional space $Z \subset \mathrm{dom}(\Theta)$ together with the estimate (4.4) further implies that $(\Theta \tilde{w}^k)_k$ is bounded. $\qquad\square$

Using the previous lemma, we now deal with the kernel and coercivity of regularization graph functionals.

Theorem 4.3. *Let $\mathcal{G}_\alpha$ be a regularization graph with weights $(\alpha_e)_e$, underlying graph structure $G = (V, E)$ and root node $\hat{n} \in V$, and let $R_\alpha = R(\mathcal{G}_\alpha) : X_{\hat{n}} \to [0, +\infty]$ be the associated regularization graph functional. Then:*

  i) The infimum in the recursive representation of R_α (3.1) provided in Lemma 3.2 is attained for any $u \in X_{\hat{n}}$.

  ii) R_α is weak lower-semicontinous.*

  iii) There exists a finite dimensional subspace $L \subset X_{\hat{n}}$ such that R_α is invariant on L and for $P_L : X \to L$ a bounded, linear projection there exist $C > 0$, $D \geq 0$ such that, for $u \in X_{\hat{n}}$,
  $$\|u - P_L u\|_{X_{\hat{n}}} \leq C R_\alpha(u) + D. \tag{4.5}$$

Proof. We prove the result via induction over the height of the graph. Assume that the claimed assertions hold true for any regularization graph of height less than h and let $\mathcal{G}_\alpha$ be a regularization graph of height h with associated functional $R_\alpha = R(\mathcal{G}_\alpha)$. If $h = 0$, the results hold trivially with $L = \{0\}$ thanks to Assumptions (H1), (H2) and the definition of trivial regularization graphs. Otherwise, using Lemma 3.2 we write R_α as

$$R_\alpha(u) = \inf \left\{ \Psi_{\hat{n}}\left(u - \sum_{\hat{e} \in \hat{E}} \alpha_{\hat{e}} \Phi_{\hat{e}} w_{\hat{e}}\right) + \sum_{\hat{e} \in \hat{E}} R^{\hat{e}}_{\alpha^{\hat{e}}}(\Theta_{\hat{e}} w_{\hat{e}}) \,\Big|\, w_{\hat{e}} \in \mathrm{dom}(\Theta_{\hat{e}}) \text{ for all } \hat{e} \in \hat{E} \right\}, \tag{4.6}$$

where $\hat{E}$ are the edges connected to $\hat{n}$ and for each $\hat{e} \in \hat{E}$, $R^{\hat{e}}_{\alpha^{\hat{e}}} : X_{n^{\hat{e}}} \to [0, \infty]$ is a functional associated to a regularization graph $\mathcal{G}^{\hat{e}}_{\alpha^{\hat{e}}}$ with root node $\hat{n}^{\hat{e}}$. Also, remember that by (H5) and (H6) (see also Remark 2.1) each $\Theta_{\hat{e}} : X_{\hat{e}} \to X_{\hat{n}^{\hat{e}}}$ has closed range, finite dimensional kernel and satisfies

$$\|w - P_{\ker(\Theta_{\hat{e}})}w\|_{X_{\hat{e}}} \leq B^{\hat{e}}\|\Theta_{\hat{e}} w\|_{X_{\hat{n}^{\hat{e}}}} \tag{4.7}$$

for $B^{\hat{e}} > 0$ and all $w \in \mathrm{dom}(\Theta_{\hat{e}})$.

Applying the induction hypothesis, each $R^{\hat{e}}_{\alpha^{\hat{e}}}$ is weak* lower-semicontinuous and there exists a finite dimensional subspace $L^{\hat{e}}$ where $R^{\hat{e}}_{\alpha^{\hat{e}}}$ is invariant such that for $P_{L^{\hat{e}}}$ a bounded linear projection there exist constants $C^{\hat{e}} > 0$ and $D^{\hat{e}} \geq 0$ with

$$\|v - P_{L^{\hat{e}}} v\|_{X_{\hat{n}^{\hat{e}}}} \leq C^{\hat{e}} R^{\hat{e}}_{\alpha^{\hat{e}}}(v) + D^{\hat{e}} \qquad \forall v \in X_{\hat{n}^{\hat{e}}}. \tag{4.8}$$

Moreover, applying Lemma 4.1 with $L = L^{\hat{e}}$, $G = P_{L^{\hat{e}}}$, $R = R_{\alpha^{\hat{e}}}^{\hat{e}}$ and $K = \mathrm{rg}(\Theta_{\hat{e}})$ (that is closed thanks to (H5), (H6) and Remark 2.1) yields that for $P_{\mathrm{rg}(\Theta_{\hat{e}}) \cap L^{\hat{e}}}$ a linear, continuous projection there exists $\tilde{C}^{\hat{e}} > 0$ and $\tilde{D}^{\hat{e}} \geq 0$ with

$$\|v - P_{\mathrm{rg}(\Theta_{\hat{e}}) \cap L^{\hat{e}}} v\|_{X_{\hat{n}\hat{e}}} \leq \tilde{C}^{\hat{e}} R_{\alpha^{\hat{e}}}^{\hat{e}}(v) + \tilde{D}^{\hat{e}} \qquad \forall v \in X_{\hat{n}\hat{e}}. \tag{4.9}$$

Now proceeding as in the proof of Lemma 4.2, we define $\ker(\Theta_{\hat{e}})^{\perp} := \mathrm{rg}(\mathrm{Id} - P_{\ker(\Theta_{\hat{e}})}) \cap \mathrm{dom}(\Theta_{\hat{e}})$. It is easy to see that $\ker(\Theta_{\hat{e}})^{\perp}$ is a complement of $\ker(\Theta_{\hat{e}})$ in $\mathrm{dom}(\Theta_{\hat{e}})$ and that $\Theta_{\hat{e}}$ is injective on $\ker(\Theta_{\hat{e}})^{\perp}$. Hence, with

$$M^{\hat{e}} := \Theta_{\hat{e}}^{-1}(L^{\hat{e}}), \tag{4.10}$$

we can define $P^{\hat{e}} : \mathrm{dom}(\Theta_{\hat{e}}) \to M^{\hat{e}}$ as

$$P^{\hat{e}} w := \Theta_{\hat{e}}^{-1} P_{\mathrm{rg}(\Theta_{\hat{e}}) \cap L^{\hat{e}}} \Theta_{\hat{e}} w + P_{\ker(\Theta_{\hat{e}})}^{\hat{e}} w, \tag{4.11}$$

where $\Theta_{\hat{e}}^{-1}$ is the inverse of $\Theta_{\hat{e}} : \ker(\Theta_{\hat{e}})^{\perp} \to \mathrm{rg}(\Theta_{\hat{e}})$. Then, we observe that $R_{\alpha^{\hat{e}}}^{\hat{e}} \circ \Theta_{\hat{e}}$ is invariant on $M^{\hat{e}}$ and that $M^{\hat{e}}$ is a finite dimensional vector space. Then, estimating as in (4.4), we obtain that for each $w \in X_{\hat{e}}$ that

$$\|w - P^{\hat{e}} w\|_{X_{\hat{e}}} \leq B^{\hat{e}} \tilde{C}^{\hat{e}} R_{\alpha^{\hat{e}}}^{\hat{e}}(\Theta_{\hat{e}} w) + B^{\hat{e}} \tilde{D}^{\hat{e}}. \tag{4.12}$$

Now we write the recursive representation (4.6) in a vectorized form as follows: With $w = (w_{\hat{e}})_{\hat{e} \in \hat{E}} \in X := \bigtimes_{\hat{e} \in \hat{E}} X_{\hat{e}}$, define $K : X \to X_{\hat{n}}$ and $\Theta : \bigtimes_{\hat{e} \in \hat{E}} \mathrm{dom}(\Theta_{\hat{e}}) \subset X \to \bigtimes_{\hat{e} \in \hat{E}} X_{\hat{n}\hat{e}}$ as

$$Kw := \sum_{\hat{e} \in \hat{E}} \alpha_{\hat{e}} \Phi_{\hat{e}} w_{\hat{e}} \quad \text{and} \quad \Theta w := (\Theta_{\hat{e}} w_{\hat{e}})_{\hat{e} \in \hat{E}}. \tag{4.13}$$

Further, with $v = (v_{\hat{n}\hat{e}})_{\hat{e} \in \hat{E}} \in \bigtimes_{\hat{e} \in \hat{E}} X_{\hat{n}\hat{e}}$ define $F : \bigtimes_{\hat{e} \in \hat{E}} X_{\hat{n}\hat{e}} \to [0, +\infty]$ as

$$F(v) = \sum_{\hat{e} \in \hat{E}} R_{\alpha^{\hat{e}}}^{\hat{e}}(v_{\hat{n}\hat{e}}). \tag{4.14}$$

Finally, define $F \circ \Theta : X \to [0, +\infty]$ and $P_M : X \to M$, where $M := \bigtimes_{\hat{e} \in \hat{E}} M^{\hat{e}}$ is a finite dimensional subspace of X, as

$$(F \circ \Theta)(w) = \begin{cases} \sum_{\hat{e} \in \hat{E}} R_{\alpha^{\hat{e}}}^{\hat{e}}(\Theta_{\hat{e}} w_{\hat{e}}) & \text{if } w \in \bigtimes_{\hat{e} \in \hat{E}} \mathrm{dom}(\Theta_{\hat{e}}), \\ \infty & \text{else,} \end{cases} \quad \text{and} \quad P_M w = (P^{\hat{e}} w_{\hat{e}})_{\hat{e} \in \hat{E}}. \tag{4.15}$$

We observe that $F \circ \Theta$ is invariant on $M = \Theta^{-1}(L)$ where $L = \bigtimes_{\hat{e} \in \hat{E}} L^{\hat{e}}$. With these definitions we can write the recursive representation (4.6) in the vectorized form

$$R_\alpha(u) = \inf \left\{ \Psi_{\hat{n}}(u - Kw) + (F \circ \Theta)(w) \,\Big|\, w \in \bigtimes_{\hat{e} \in \hat{E}} \mathrm{dom}(\Theta_{\hat{e}}) \right\}.$$

Now we first show weak* lower semi-continuity of R_α on $X_{\hat{n}}$. To this aim, take $(u^k)_k$ to be a sequence in $X_{\hat{n}}$ converging weakly* to some $u \in X_{\hat{n}}$. Without loss of generality, we can assume that $\liminf_k R_\alpha(u^k) < \infty$ and, up to extracting a subsequence, we can assume that $(u^k)_k$ realizes the $\liminf$. Next, take $(w^k)_k$ to be a sequence in X such that

$$\Psi_{\hat{n}}(u^k - Kw^k) + (F \circ \Theta)(w^k) \leq R_\alpha(u^k) + 1/k.$$

Together with assumption (H2), this implies boundedness of $(Kw^k)_k$ and $(F \circ \Theta)(w^k)$. We now want to apply Lemma 4.2 choosing $\bigtimes_{\hat{e} \in \hat{E}} L^{\hat{e}}$ for L and $v \mapsto (P_{L^{\hat{e}}} v_{\hat{n}\hat{e}})_{\hat{e} \in \hat{E}}$ with $v = (v_{\hat{n}\hat{e}})_{\hat{e} \in \hat{E}} \in \bigtimes_{\hat{e} \in \hat{E}} X_{\hat{n}\hat{e}}$ for P_L. Note that F is weak* lower semicontinuous, it is invariant on $\bigtimes_{\hat{e} \in \hat{E}} L^{\hat{e}}$ and the estimate (4.2) holds thanks to the inductive assumption (4.8) applied to each $R_{\alpha^{\hat{e}}}^{\hat{e}}$. Moreover, it can be readily verified that the operator Θ is weak* to weak* closed, has finite dimensional kernel and the estimate (4.3) holds as a direct consequence of (4.7) applied to each $\Theta_{\hat{e}}$. So, applying Lemma 4.2 and using the weak* to weak* closedness of Θ we can select, up to a non-relabeled subsequence, $(\tilde{w}^k)_k$ such that $\tilde{w}^k - w^k \in \ker(K) \cap M$, $\tilde{w}^k$ converges weak* to some $w \in X_{\hat{n}}$ and $\Theta \tilde{w}^k$ converges weak* to Θw. Weak* lower semi-continuity of $\Psi_{\hat{n}}$ and $F \circ \Theta$ together with weak* continuity of K and the invariance of $F \circ \Theta$ on M then imply

$$R_\alpha(u) \le \Psi_{\hat{n}}(u - Kw) + F(\Theta w) \le \liminf_k \Psi_{\hat{n}}(u^k - Kw^k) + F(\Theta w^k) \le \liminf_k R_\alpha(u^k),$$

thus weak* lower semi-continuity of R_α on $X_{\hat{n}}$ follows. In addition, given any $u \in X_{\hat{n}}$, choosing $u^k = u$ for every k implies existence of minimizers in (4.6) as claimed.

Now we note that R_α is invariant on the finite dimensional space

$$L := \underset{\hat{e} \in \hat{E}}{+} \alpha_{\hat{e}} \Phi_{\hat{e}}(M^{\hat{e}}).$$

Indeed, if $u = \sum_{\hat{e} \in \hat{E}} \alpha_{\hat{e}} \Phi_{\hat{e}} u_{\hat{e}}$ with $u_{\hat{e}} \in M^{\hat{e}}$ it follows from $(F \circ \Theta)((\alpha_{\hat{e}} u_{\hat{e}})_{\hat{e} \in \hat{E}}) = 0$ that $R_\alpha(\tilde{u} + u) = R_\alpha(\tilde{u})$ for all $\tilde{u} \in X_{\hat{n}}$.

Next we show the coercivity estimate. To this aim, for any given $u \in X_{\hat{n}}$, we select $(w_{\hat{e}})_{\hat{e} \in \hat{E}}$ to be minimizers in (4.6) and define $v = v(u) := \sum_{\hat{e} \in \hat{E}} \alpha_{\hat{e}} \Phi_{\hat{e}} P^{\hat{e}} w_{\hat{e}}$. Using (4.12), the coercivity of $\Psi_{\hat{n}}$ (see Remark 2.2) and the continuity of $\Phi_{\hat{e}}$ we estimate

$$\|u - v(u)\|_{X_{\hat{n}}} = \left\| u - \sum_{\hat{e} \in \hat{E}} \alpha_{\hat{e}} \Phi_{\hat{e}} P^{\hat{e}} w_{\hat{e}} \right\|_{X_{\hat{n}}}$$

$$= \left\| u - \sum_{\hat{e} \in \hat{E}} \alpha_{\hat{e}} \Phi_{\hat{e}} w_{\hat{e}} + \sum_{\hat{e} \in \hat{E}} \alpha_{\hat{e}} \Phi_{\hat{e}} w_{\hat{e}} - \sum_{\hat{e} \in \hat{E}} \alpha_{\hat{e}} \Phi_{\hat{e}} P^{\hat{e}} w_{\hat{e}} \right\|_{X_{\hat{n}}}$$

$$\le C_1 \Psi_{\hat{n}}\left(u - \sum_{\hat{e} \in \hat{E}} \alpha_{\hat{e}} \Phi_{\hat{e}} w_{\hat{e}} \right) + D_1 + C_2 \sum_{\hat{e} \in \hat{E}} \alpha_{\hat{e}} \|w_{\hat{e}} - P^{\hat{e}} w_{\hat{e}}\|_{X_{\hat{e}}}$$

$$\le C_1 \Psi_{\hat{n}}\left(u - \sum_{\hat{e} \in \hat{E}} \alpha_{\hat{e}} \Phi_{\hat{e}} w_{\hat{e}} \right) + D_1 + C_2 \sum_{\hat{e} \in \hat{E}} (\alpha_{\hat{e}} B^{\hat{e}} \tilde{C}^{\hat{e}} R_{\alpha^{\hat{e}}}^{\hat{e}}(\Theta_{\hat{e}} w_{\hat{e}}) + \alpha_{\hat{e}} B^{\hat{e}} \tilde{D}^{\hat{e}})$$

$$\le C R_\alpha(u) + D,$$

with suitable $C > 0$ and $D \ge 0$. Using Lemma 4.1 with $G(u) = v(u)$ and $K = X_{\hat{n}}$, this implies the claimed coercivity. $\square$

Remark 4.4 (Domain and invariant subspace). With $R_\alpha : X_{\hat{n}} \to [0, \infty]$ a regularization graph functional with recursive representation as in Lemma 3.2, we have that

$$\mathrm{dom}(R_\alpha) = \mathrm{dom}(\Psi_{\hat{n}}) + \underset{\hat{e} \in \hat{E}}{+} \alpha_{\hat{e}} \Phi_{\hat{e}}\left(\Theta_{\hat{e}}^{-1}(\mathrm{dom}(R_{\alpha^{\hat{e}}}^{\hat{e}})) \right),$$

22

where equality follows since the infimum in the recursive represenation of Lemma 3.2 is attained. Further, a simple contradiction argument shows that the finite dimensional subspace L where R_α is invariant and coercive in the sense of (4.5) is unique (and will henceforth be called *the invariant subspace of R_α*). We also have the recursive representation

$$L = \bigsqcup_{\hat{e} \in \hat{E}} \alpha_{\hat{e}} \Phi_{\hat{e}} \Big(\Theta_{\hat{e}}^{-1}(L^{\hat{e}})) \Big),$$

with $L_{\hat{e}}$ the invariant subspace of $R_{\alpha_{\hat{e}}}^{\hat{e}}$. Via induction, this implies in particular that L only depends on the support of $(\alpha_e)_e$, i.e., the set of edges where $\alpha_e \neq 0$, but not on their values.

Existence for the infimum over edge variables associated with edges connected to the root node in the recursive representation of R_α, as stated in Theorem 4.3, immediately implies, via induction, existence of infimizing edge variables in the definition of the regularization graph for all edges. This is stated in the following corollary.

Corollary 4.5 (Existence of infimizing edge variables)**.** *Let $\mathcal{G}_\alpha$ be a regularization graph with root node $\hat{n}$ and $R_\alpha = R(\mathcal{G}_\alpha)$ be the associated functional. Then, for each $u \in X_{\hat{n}}$, there exists $(w_e)_{e \in E}$ such that*

$$R_\alpha(u) = \sum_{n \in V} \Psi_n \Big(\Theta_{(n^-,n)} w_{(n^-,n)} - \sum_{(n,m) \in E} \alpha_{(n,m)} \Phi_{(n,m)} w_{(n,m)} \Big),$$

i.e., the infimum in the definition of the regularization graph functional is attained.

Remark 4.6 (Regularity). Let us observe how an infimal-based combination and an extension of regularization graphs affect the coercivity of regularization graph functionals as in Theorem 4.3.

- When combining two different regularization graph functionals defined on two different normed spaces via infimal convolution, the norm for underlying joint space and hence the norm for the coercivity estimate needs to be the weaker of the two norms. In the construction of Proposition 3.5, this is reflected in the assumption that the embeddings $I_X^1 : X_{\hat{n}^1} \to X$ and $I_X^2 : X_{\hat{n}^2} \to X$ need to be weak* continuous. An example here is the infimal convolution of TV and TV^2, where TV and TV^2 are coercive up to their kernels on $L^{d/(d-1)}$ and $L^{d/(d-2)}$, respectively (here the exponents are set to ∞ for $d = 1$ and $d \leq 2$, respectively). The infimal-convolution-based combination of the regularization graphs corresponding to TV and TV^2, according to Proposition 3.5, is then coercive on the weaker space $L^{d/(d-1)}$, see [7, Section 4.2] for details.

- When extending a given regularization graph with a further edge, stronger norms can be chosen. A particular example is the composition of two gradient operators ∇_1, ∇_2 to obtain TV$^2 = \|(\nabla_2 \circ \nabla_1) \cdot \|_M$. Given that ∇_2 is coercive up to its kernel on $L^{d/(d-1)}$, we can define ∇_1 as operator from $L^{d/(d-2)}$ to $L^{d/(d-1)}$ and again obtain coercivity up to constant functions between those spaces by standard Sobolev embeddings. In this case, the overall regularization graph functional corresponding to TV^2 is coercive up to affine functions with respect to the norm in $L^{d/(d-2)}$, which is the improved regularity of TV^2, see [7, Section 3].

The following proposition deals with the dependence of the coercivity estimate for a regularization graph functional R_α on the weights α.

Proposition 4.7 (Dependence on the weights). *Let $\mathcal{G}_\alpha$ be a regularization graph with weights α and root node $\hat{n}$, let L be the invariant subspace of $R_\alpha = R(\mathcal{G}_\alpha)$ that only depends on the structure of the regularization graph $\mathcal{G}$ and the support of $(\alpha_e)_{e \in E}$, let $K \subset X_{\hat{n}}$ be a closed subspace and let $P_{K \cap L} : X_{\hat{n}} \to L$ be a bounded, linear projection. Then, for any $A \geq \max\{\alpha_e \,|\, e \in E\}$ there exist $C > 0$, $D \geq 0$ that only depend on A such that for any $u \in X_{\hat{n}}$ we have*

$$\|u - P_L u\|_{X_{\hat{n}}} \leq C_\alpha C R_\alpha(u) + D,$$

where

$$C_\alpha := \max\Big\{ \prod_{e \in F} \alpha_e \,\Big|\, F \subset E \text{ is either empty or a chain with } \hat{n} \in V \text{ as root} \Big\},$$

using the same conventions as in Lemma 3.3.

Proof. This follows from first applying Theorem 4.3 and, subsequently, Lemma 4.1 to $R_{\tilde{\alpha}} = R(\mathcal{G}_{\tilde{\alpha}})$, where $\tilde{\alpha}_e = A$ if $\alpha_e > 0$ and $\tilde{\alpha}_e = 0$ else, and then using the estimate of Lemma 3.3. $\square$

The next proposition deals with extending a regularization graph functional $R_\alpha : X_{\hat{n}} \to [0, \infty]$ by infinity to a Banach space X not satisfying hypothesis (H8), but with $X_{\hat{n}} \hookrightarrow X$. The prototypical application of this result would be, e.g., to extend R_α from $L^p(\Omega)$ with $p > 1$ to $L^1(\Omega)$, where $\Omega \subset \mathbb{R}^d$ is a bounded domain. Note that directly choosing $X_{\hat{n}} = L^1(\Omega)$ is not feasible since, in general, bounded sequences in $L^1(\Omega)$ do not admit weak* convergent subsequences (or weakly convergent subsequences since $L^1(\Omega)$ is generally not a dual space).

Proposition 4.8 (Extended domain). *Let $\mathcal{G}_\alpha$ be a regularization graph with weights α and root node $\hat{n}$. Let L be its invariant space. Suppose that X is a Banach space such that $X_{\hat{n}} \hookrightarrow X$ and $X_{\hat{n}}$ is reflexive. Then, with $R_\alpha = R(\mathcal{G}_\alpha)$ extended to X via $R_\alpha(x) = \infty$ for $x \in X \setminus X_{\hat{n}}$, R_α is convex and lower semi-continuous w.r.t. weak convergence in X, and for any continuous, linear projection $\tilde{P}_L : X \to L$, there exists $C > 0$, $D \geq 0$ such that*

$$\|u - \tilde{P}_L u\|_X \leq C R_\alpha(u) + D \quad \text{for all } u \in X. \tag{4.16}$$

Proof. Convexity is immediate and the coercivity estimate follows directly from the continuous embedding $X_{\hat{n}} \hookrightarrow X$ and Theorem 4.3 by defining P_L as the restriction of $\tilde{P}_L$ to $X_{\hat{n}}$. Regarding weak lower semi-continuity, take $(u^k)_k$ to be a sequence in X converging weakly to some $u \in X$. Without loss of generality, we can assume that $\liminf_k R_\alpha(u^k) < \infty$ and, up to extracting a subsequence, we can choose u^k such that it realizes the liminf and $u^k \in X_{\hat{n}}$ for every k. With $P_L : X_{\hat{n}} \to L$ a continuous, linear projection, from the coercivity estimate of Theorem 4.3 applied to $R_\alpha : X_{\hat{n}} \to [0, \infty]$ we obtain that $v^k := u^k - P_L u^k$ is bounded in $X_{\hat{n}}$ such that we may assume weak convergence of the latter to $v \in X_{\hat{n}}$. Also, by the embedding $X_{\hat{n}} \hookrightarrow X$, $P_L u^k = u^k - v^k$ is bounded in X and hence, by finite dimensionality of L, admits a subsequence converging to some $z \in X_{\hat{n}} \cap L$ with respect to $\|\cdot\|_{X_{\hat{n}}}$. Again by the embedding $X_{\hat{n}} \hookrightarrow X$, weak convergence in $X_{\hat{n}}$ implies weak convergence in X such that, by uniqueness of the weak limit, $u = v + z \in X_{\hat{n}}$. Lower semi-continuity of R_α with respect to weak convergence in $X_{\hat{n}}$ finally implies

$$R_\alpha(u) = R_\alpha(v + z) \leq \liminf_k R_\alpha(v^k + P_L u^k) = \liminf_k R_\alpha(u^k)$$

implying the lower semi-continuity of R_α with respect to weak convergence in X. $\square$

Remark 4.9. It can be observed that, in the above result, reflexivity of $X_{\hat{n}}$ (instead of just requiring that bounded sequences admit weak* convergent subsequences) is only needed to conclude from (weak) convergence of $(v^k + P_L u^k)_k$ to $v + z$ in $X_{\hat{n}}$ and the weak convergence of $(v^k + P_L u^k)_k$ to u in X that, by uniqueness of limits, $v + z = u$ follows. The same could be achieved for weak* convergence of $(v^k + P_L u^k)_k$ in $X_{\hat{n}}$, thus not requiring reflexivity, if, for instance, $X_{\hat{n}} = L^\infty(\Omega)$ and $X = L^1(\Omega)$.

5 Predual formulation of regularization graphs

The goal of this section is to provide an equivalent, predual reformulation of regularization graphs. Remember that a regularization graph functional $R_\alpha : X_{\hat{n}} \to [0, \infty]$ can be written in a vectorized form as

$$R_\alpha(u) = \inf \{ \Psi_u(v) \mid v \in \mathrm{rg}(\Lambda_\alpha) \}$$

with Λ_α and Ψ_u given in (2.2) and (2.3), respectively. With $\Lambda_\alpha^\#$ and $\Psi_u^\#$ predual versions of Λ_α and Ψ_u, respectively, our goal is to show that every regularization graph functional R_α can be written equivalently as

$$R_\alpha(u) = - \inf \left\{ \Psi_u^\#(v) \mid v \in \ker(\Lambda_\alpha^\#) \right\}.$$

To this aim, we need in particular that the functionals Ψ_n and the operators Θ_e and Φ_e admit predual versions. By an application of the Fenchel–Moreau theorem [22, Proposition 4.1] it is easy to see that there exist convex, proper, lower semicontinuous functionals $\Psi_n^\# : X_n^\# \to [0, \infty]$ such that their convex conjugates are Ψ_n.

Lemma 5.1. *For each $n \in V$, $\Psi_n : X_n \to [0, \infty]$ is the convex conjugate of a convex, proper, lower semicontinuous functional $\Psi_n^\# : X_n^\# \to [0, \infty]$.*

Proof. Consider the dual pair (V, V^*) for $V = (X_n, \mathsf{w^*})$ and $V^* = (X_n^\#, \mathsf{w})$. Note that Ψ_n is convex, proper and lower semicontinuous on V. Therefore by the Fenchel–Moreau theorem [22, Proposition 4.1] there holds that $\Psi_n^{**} = \Psi_n$. In particular, defining $\Psi_n^\# = \Psi_n^*$, we have that $\Psi_n^\# : X_n^\# \to [-\infty, +\infty]$ is proper, convex and strongly lower semicontinuous and its convex conjugate is Ψ_n. The positivity of $\Psi_n^\#$ follows from Assumption (H3). $\qquad\square$

Moreover, Remark 2.1 ensures the existence of a bounded predual of $\Phi_{(n,m)}$ that we are going to denote by $\Phi_{(n,m)}^\# : X_n^\# \to X_{(n,m)}^\#$. Finally, we suppose that the operators Θ_e admit closed, densely defined preadjoints as stated in the following additional assumption.

(H9) For each $e = (n, m) \in E$, $\Theta_{(n,m)}$ is the adjoint of a closed, densely defined operator $\Theta_{(n,m)}^\# : \mathrm{dom}(\Theta_{(n,m)}^\#) \subset X_m^\# \to X_{(n,m)}^\#$.

Define the following predual spaces of X_V and X_E:

$$X_V^\# = \bigtimes_{n \in V} X_n^\# \quad \text{and} \quad X_E^\# = \bigtimes_{e \in E} X_e^\#. \tag{5.1}$$

Now we characterize the predual of the linear operator $\Lambda_\alpha : X_E \to X_V$ from (2.2).

Lemma 5.2. *Let $\mathcal{G}_\alpha$ be a regularization graph with root node $\hat{n}$ and corresponding operator Λ_α as in (2.2) such that (H9) holds. Then Λ_α is the dual of the linear operator $\Lambda_\alpha^\# : \mathrm{dom}(\Lambda_\alpha^\#) \subset X_V^\# \to X_E^\#$ with*

$$\mathrm{dom}(\Lambda_\alpha^\#) = \{ v \in X_V^\# \mid \forall (n,m) \in E, \ v_m \in \mathrm{dom}(\Theta_{(n,m)}^\#) \}$$

given as

$$(\Lambda_\alpha^\# v)_{(n,m)} = \Theta_{(n,m)}^\# v_m - \alpha_{(n,m)} \Phi_{(n,m)}^\# v_n \quad \textit{for all} \quad v \in \mathrm{dom}(\Lambda_\alpha^\#). \tag{5.2}$$

Proof. First, let us verify that $\mathrm{dom}((\Lambda_\alpha^\#)^*) = \mathrm{dom}(\Lambda_\alpha)$, where $\mathrm{dom}(\Lambda_\alpha) = \bigtimes_{e \in E} \mathrm{dom}(\Theta_e)$. Note that for every $v \in \mathrm{dom}(\Lambda_\alpha^\#)$ and $w \in X_E$ it holds

$$\langle w, \Lambda_\alpha^\# v \rangle = \sum_{(n,m) \in E} \langle w_{(n,m)}, \Theta_{(n,m)}^\# v_m - \alpha_{(n,m)} \Phi_{(n,m)}^\# v_n \rangle$$

$$= \sum_{m \in V} \sum_{(n,m) \in E} \langle w_{(n,m)}, \Theta_{(n,m)}^\# v_m \rangle - \sum_{n \in V} \sum_{(n,m) \in E} \langle w_{(n,m)}, \alpha_{(n,m)} \Phi_{(n,m)}^\# v_n \rangle. \tag{5.3}$$

By the definition of the domain of the adjoint we have

$$\mathrm{dom}((\Lambda_\alpha^\#)^*) = \{ w \in X_E \mid \exists C > 0 \ \text{ such that } \ \langle w, \Lambda_\alpha^\# v \rangle \leq C \|v\|_{X_V^\#} \quad \forall v \in \mathrm{dom}(\Lambda_\alpha^\#) \}. \tag{5.4}$$

Therefore, using (5.3), the boundedness of $\Phi_{(n,m)}$ and the fact that $\Theta_{(n,m)}$ is the adjoint of $\Theta_{(n,m)}^\#$ (see Assumption (H9)), we immediately deduce that $\mathrm{dom}(\Lambda_\alpha) \subset \mathrm{dom}((\Lambda_\alpha^\#)^*)$. To prove the opposite inclusion, consider $w = (w_e)_{e \in E} \notin \mathrm{dom}(\Lambda_\alpha)$. Then, there exists $(n,m) \in E$ such that $w_{(n,m)} \notin \mathrm{dom}(\Theta_{(n,m)})$. Using that

$$\mathrm{dom}(\Theta_{(n,m)}) = \mathrm{dom}((\Theta_{(n,m)}^\#)^*)$$
$$= \{ w \in X_{(n,m)} \mid \exists C > 0 \ \text{ such that } \ \langle w, \Theta_{(n,m)}^\# v \rangle \leq C \|v\|_{X_m^\#} \forall v \in \mathrm{dom}(\Theta_{(n,m)}^\#) \}$$

(see Assumption (H9)), we can find a sequence $(v_m^k)_k$ with $v_m^k \in \mathrm{dom}(\Theta_{(n,m)}^\#) \setminus \{0\}$ for each k such that

$$\lim_{k \to \infty} \frac{1}{\|v_m^k\|_{X_m^\#}} \langle w_{(n,m)}, \Theta_{(n,m)}^\# v_m^k \rangle = +\infty. \tag{5.5}$$

Then, with $v^k \in X_V^\#$ defined as $(v^k)_{\tilde{m}} = v_m^k$ if $\tilde{m} = m$ and $(v^k)_{\tilde{m}} = 0$ else, one notices that, using the boundedness of $\Phi_e^\#$ for every $e \in E$, there exists a positive constant C independent on k such that

$$\frac{1}{\|v^k\|_{X_V^\#}} \langle w, \Lambda_\alpha^\# v^k \rangle = \frac{1}{\|v_m^k\|_{X_m^\#}} \langle w_{(n,m)}, \Theta_{(n,m)}^\# v_m^k \rangle - \sum_{\tilde{n} \, : \, (m,\tilde{n}) \in E} \frac{\alpha_{(m,\tilde{n})}}{\|v_m^k\|_{X_m^\#}} \langle w_{(m,\tilde{n})}, \Phi_{(m,\tilde{n})}^\# v_m^k \rangle$$

$$\geq \frac{1}{\|v_m^k\|_{X_m^\#}} \langle w_{(n,m)}, \Theta_{(n,m)}^\# v_m^k \rangle - C. \tag{5.6}$$

Therefore, combining (5.6) and (5.5) with (5.4), we deduce that $w \notin \mathrm{dom}((\Lambda_\alpha^\#)^*)$ showing that $\mathrm{dom}((\Lambda_\alpha^\#)^*) \subset \mathrm{dom}(\Lambda_\alpha)$.

26

It remains to prove (5.2). For every $v \in \text{dom}(\Lambda_\alpha^\#)$ and $w \in \text{dom}(\Lambda_\alpha)$, using (5.3) we have

$$
\begin{aligned}
\langle (\Lambda_\alpha^\#)^* w, v \rangle &= \langle w, \Lambda_\alpha^\# v \rangle \\
&= \sum_{m \in V \setminus \{\hat{n}\}} \langle \Theta_{(m^-,m)} w_{(m^-,m)}, v_m \rangle - \sum_{m \in V} \sum_{(m,n) \in E} \langle \alpha_{(m,n)} \Phi_{(m,n)} w_{(m,n)}, v_m \rangle \\
&= \sum_{m \in V \setminus \{\hat{n}\}} \left\langle \Theta_{(m^-,m)} w_{(m^-,m)} - \sum_{(m,n) \in E} \alpha_{(m,n)} \Phi_{(m,n)} w_{(m,n)}, v_m \right\rangle \\
&\quad + \left\langle -\sum_{(\hat{n},n) \in E} \alpha_{(\hat{n},n)} \Phi_{(\hat{n},n)} w_{(\hat{n},n)}, v_{\hat{n}} \right\rangle \\
&= \langle \Lambda_\alpha w, v \rangle
\end{aligned}
$$

ending the proof. □

With this, we define a predual regularization graph functional as follows.

Definition 5.3 (Predual regularization graph functional)*. Given a regularization graph $\mathcal{G}_\alpha$ with root node $\hat{n}$ and associated functional $R_\alpha = R(\mathcal{G}_\alpha)$ such that (H9) holds, we define the predual regularization graph functional $R_\alpha^\# = R^\#(\mathcal{G}_\alpha) : X_{\hat{n}} \to [0, +\infty]$ as*

$$
\begin{aligned}
R_\alpha^\#(u) = -\inf \Bigg\{ &\Psi_{\hat{n}}^\#(v_{\hat{n}}) - \langle u, v_{\hat{n}} \rangle + \sum_{n \in V \setminus \{\hat{n}\}} \Psi_n^\#(v_n) \,\Bigg|\, (v_n)_{n \in V} \in \text{dom}(\Lambda_\alpha^\#), \\
&\Theta_{(n,m)}^\# v_m = \alpha_{(n,m)} \Phi_{(n,m)}^\# v_n \text{ for all } (n,m) \in E \Bigg\} \\
= -\inf &\left\{ \Psi_u^\#(v) \,\big|\, v \in \ker(\Lambda_\alpha^\#) \right\}
\end{aligned}
$$

with $\Psi_u^\#(v) := \Psi_{\hat{n}}^\#(v_{\hat{n}}) - \langle u, v_{\hat{n}} \rangle + \sum_{n \in V \setminus \{\hat{n}\}} \Psi_n^\#(v_n)$, and $\Theta_e^\#$, $\Phi_e^\#$ and $\Psi_n^\#$ being the predual operators and functionals of Θ_e, Φ_e and Ψ_n, respectively, according to hypothesis (H9), Remark 2.1 and Lemma 5.1.

Our goal is now to show that $R_\alpha^\# = R_\alpha$. As first step, we obtain the following proposition.

Proposition 5.4. *Assuming again that hypothesis (H9) holds, any predual regularization graph functional $R_\alpha^\# = R^\#(\mathcal{G}_\alpha)$ according to Definition 5.3, where $\mathcal{G}_\alpha$ is a regularization graph with root node $\hat{n}$, can be written as*

$$
R_\alpha^\#(u) = \min \left\{ \Psi_u(v) \,\big|\, v \in \overline{\text{rg}(\Lambda_\alpha)}^{w^*} \right\}, \tag{5.7}
$$

where Ψ_u is defined as in (2.3).

Proof. Since $\Psi_n^\#$ is proper, convex and lower semicontinuous according to Lemma 5.1, the Fenchel–Moreau theorem [22, Proposition 4.1] yields $\Psi_n^\# = (\Psi_n^\#)^{**}$. Hence, since Ψ_n is coercive thanks to hypothesis (H2), using [45, Theorem 10], we deduce that $\Psi_n^\#$ is continuous in zero. Consequently, also $v \mapsto \Psi_{\hat{n}}^\#(v) - \langle u, v \rangle$ is continuous in zero, such that $\text{dom}(\Psi_u)$ contains a neighborhood of zero and hence is absorbing.

In order to apply the Attouch–Brézis theorem [2, Corollary 2.2], we note that

$$\bigcup_{\lambda \geq 0} \lambda(\mathrm{dom}(\Psi_u^{\#}) + \ker(\Lambda_\alpha^{\#})) = \bigcup_{\lambda \geq 0} \lambda \mathrm{dom}(\Psi_u^{\#}) + \ker(\Lambda_\alpha^{\#}) = X_V.$$

Therefore, noting that $(\Psi_u^{\#})^* = \Psi_u$ we deduce that

$$-\inf\left\{\Psi_u^{\#}(v)\,|\,v \in \ker(\Lambda_\alpha^{\#})\right\} = \min\left\{\Psi_u(v)\,|\,v \in (\ker(\Lambda_\alpha^{\#}))^{\perp}\right\},$$

where the minimum is attained and $(\ker(\Lambda_\alpha^{\#}))^{\perp}$ denotes the annihilator of $\ker(\Lambda_\alpha^{\#})$. As $\Lambda_\alpha^{\#}$ is closed and densely defined we apply Remark 17 in [11] to deduce that $(\ker(\Lambda_\alpha^{\#}))^{\perp} = \overline{\mathrm{rg}(\Lambda_\alpha)}^{w*}$ and conclude. $\qquad\square$

In the next proposition we use Theorem 4.3 to prove that $\mathrm{rg}(\Lambda_\alpha^{\#})$ is weak*-closed, and hence obtain the desired duality formulation.

Proposition 5.5. *Let $\Lambda_\alpha : X_E \to X_V$ be as in (2.2) corresponding to a regularization graph $\mathcal{G}_\alpha$. Then, $\mathrm{rg}(\Lambda_\alpha)$ is weak*-closed in X_V. If in addition (H9) holds and $R_\alpha = R(\mathcal{G}_\alpha)$ and $R_\alpha^{\#} = R^{\#}(\mathcal{G}_\alpha)$ are the primal and predual regularization graph functionals according to Definitions 2.3 and 5.3, respectively, then*

$$R_\alpha^{\#} = R_\alpha.$$

Proof. We are only going to show weak* closedness of $\mathrm{rg}(\Lambda_\alpha)$ since, under assumption (H9), the assertion $R_\alpha^{\#} = R_\alpha$ then follows as immediate consequence of Proposition 5.4 and the definition of R_α. Assume that the result holds true for all operators corresponding to a regularization graph of height less than h and let $\mathcal{G}_\alpha$ be a regularization graph of height h with corresponding operator Λ_α and associated directed graph $G = (V, E)$. The case $h = 0$ is immediate since $\mathrm{rg}(\Lambda_\alpha) = \{0\}$.

Denote by $\hat{n}$ the root node of $\mathcal{G}_\alpha$ and let $\hat{E}$ be the edges connected to the root node and $\hat{V}$ their corresponding endpoints. Further, for $\hat{e} = (\hat{n}, \hat{n}^{\hat{e}}) \in \hat{E}$, denote by $G^{\hat{e}} = (V^{\hat{e}}, E^{\hat{e}})$ the subtree of G with root node $\hat{n}^{\hat{e}}$ and by $\Lambda_{\alpha^{\hat{e}}}^{\hat{e}}$ the operator corresponding to the subtree $G^{\hat{e}}$. Then, we define $\mathcal{G}_{\alpha^{\hat{e}}}^{\hat{e}}$ to be the regularization graph with structure $\mathcal{G}^{\hat{e}} = (G^{\hat{e}}, (\|\cdot\|_{X_n})_{n \in V^{\hat{e}}}, (\Theta_e)_{e \in E^{\hat{e}}}, (\Phi_e)_{e \in E^{\hat{e}}})$ and weights $\alpha^{\hat{e}} = (\alpha_e)_{e \in E^{\hat{e}}}$. It follows that $\mathcal{G}_{\alpha^{\hat{e}}}^{\hat{e}}$ is indeed a regularization graph and that the associated functional $R_{\alpha^{\hat{e}}}^{\hat{e}} = R(\mathcal{G}_{\alpha^{\hat{e}}}^{\hat{e}}) : X_{\hat{n}^{\hat{e}}} \to [0, \infty)$ is given as

$$R_{\alpha^{\hat{e}}}^{\hat{e}}(u) = \inf\{\Upsilon_u^{\hat{e}}(v)\,|\,v \in \mathrm{rg}(\Lambda_{\alpha^{\hat{e}}}^{\hat{e}})\} \tag{5.8}$$

with $\Upsilon_u^{\hat{e}}(v) := \|u + v_{\hat{n}^{\hat{e}}}\|_{X_{\hat{n}^{\hat{e}}}} + \sum_{n \in V^{\hat{e}} \setminus \{\hat{n}^{\hat{e}}\}} \|v_n\|_{X_n}$.

Now take $(\Lambda_\alpha w^k)_k$ with $(w^k)_k$ in $\mathrm{dom}(\Lambda_\alpha)$ to be a sequence in $\mathrm{rg}(\Lambda_\alpha)$ weak*-converging to some $y \in X_V = \bigtimes_{n \in V} X_n$. Then, we note that by the definition of $R_{\alpha^{\hat{e}}}^{\hat{e}}$ and $\Lambda_{\alpha^{\hat{e}}}^{\hat{e}}$ we have that

$$R_{\alpha^{\hat{e}}}^{\hat{e}}(\Theta_{\hat{e}} w_{\hat{e}}^k) \leq \sum_{n \in V^{\hat{e}}} \|(\Lambda_\alpha w^k)_n\|_{X_n} < \infty.$$

Further, defining $w_{\hat{E}}^k = (w_{\hat{e}}^k)_{\hat{e} \in \hat{E}}$ and, for $w_{\hat{E}} = (w_{\hat{e}})_{\hat{e} \in \hat{E}}$, $\Theta_{\hat{E}} w_{\hat{E}} = (\Theta_{\hat{e}} w_{\hat{e}})_{\hat{e} \in \hat{E}}$,

$$K w_{\hat{E}} := \sum_{\hat{e} \in \hat{E}} \alpha_{\hat{e}} \Phi_{\hat{e}} w_{\hat{e}}, \quad \text{and} \quad F((v_{\hat{n}^{\hat{e}}})_{\hat{e} \in \hat{E}}) := \sum_{\hat{e} \in \hat{E}} R_{\alpha^{\hat{e}}}^{\hat{e}}(v_{\hat{n}^{\hat{e}}}),$$

we observe that both $(Kw_{\hat{E}}^k)_k$ and $((F \circ \Theta_{\hat{E}})(w_{\hat{E}}^k))_k$ are bounded. Hence, using Lemma 4.2, we can choose $(\tilde{w}_{\hat{E}}^k)_k$ such that $w_{\hat{E}}^k - \tilde{w}_{\hat{E}}^k \in \ker(K) \cap \Theta_{\hat{E}}^{-1}(L_{\hat{E}})$ and both $(\tilde{w}_{\hat{E}}^k)_k$ and $(\Theta_{\hat{E}}\tilde{w}_{\hat{E}}^k)_k$ are bounded, where $L_{\hat{E}}$ is the invariant subspace of F is given as $L_{\hat{E}} = \bigtimes_{\hat{e} \in \hat{E}} L_{\hat{e}}$ with $L_{\hat{e}}$ the invariant subspace of $R_{\alpha^{\hat{e}}}^{\hat{e}}$ and the respective projection onto $L_{\hat{E}}$ is given as $P_{L_{\hat{E}}}w_{\hat{E}} = (P_{L_{\hat{e}}}w_{\hat{E}})_{\hat{e} \in \hat{E}}$, with $P_{L_{\hat{e}}}$ the projection onto $L_{\hat{e}}$. Hence, up to taking a further non-relabeled subsequence, we can assume that both $(\tilde{w}_{\hat{E}}^k)_k$ and $(\Theta_{\hat{E}}\tilde{w}_{\hat{E}}^k)_k$ are weak* converging, such that, by weak* closedness of $\Theta_{\hat{E}}$, for $w_{\hat{E}} := \mathsf{w^*}\text{-}\lim_k \tilde{w}_{\hat{E}}^k$ we have that $\Theta_{\hat{E}}w_{\hat{E}} = \mathsf{w^*}\text{-}\lim_k \Theta_{\hat{E}}\tilde{w}_{\hat{E}}^k$. Now since the infimum in the definition of $R_{\alpha^{\hat{e}}}^{\hat{e}}$ as in (5.8) is attained thanks to Corollary 4.5, and since $R_{\alpha^{\hat{e}}}^{\hat{e}}(\Theta_{\hat{e}}(w_{\hat{e}}^k - \tilde{w}_{\hat{e}}^k)) = 0$, there exist minimizers $z_{E^{\hat{e}}}^k \in \mathrm{dom}(\Lambda_{\alpha^{\hat{e}}}^{\hat{e}})$ such that $0 = \Theta_{\hat{e}}(\tilde{w}_{\hat{e}}^k - w_{\hat{e}}^k) - (\Lambda_{\alpha^{\hat{e}}}^{\hat{e}}z_{E^{\hat{e}}}^k)_{\hat{n}^{\hat{e}}}$ and $0 = (\Lambda_{\alpha^{\hat{e}}}^{\hat{e}}z_{E^{\hat{e}}}^k)_n$ for all $n \in V^{\hat{e}} \setminus \{\hat{n}^{\hat{e}}\}$. Defining, with $w_{E^{\hat{e}}}^k = (w_e^k)_{e \in E^{\hat{e}}}$,

$$\tilde{w}^k := (\tilde{w}_{\hat{E}}^k, (w_{E^{\hat{e}}}^k - z_{E^{\hat{e}}}^k)_{\hat{e} \in \hat{E}})$$

we observe that $(\Lambda_\alpha \tilde{w}^k)_{\hat{n}} = (\Lambda_\alpha w^k)_{\hat{n}}$ since $w_{\hat{E}}^k - \tilde{w}_{\hat{E}}^k \in \ker(K)$, and

$$(\Lambda_\alpha \tilde{w}^k)_{\hat{n}^{\hat{e}}} = \Theta_{\hat{e}}\tilde{w}_{\hat{e}}^k - (\Lambda_{\alpha^{\hat{e}}}^{\hat{e}}(w_{E^{\hat{e}}}^k - z_{E^{\hat{e}}}^k))_{\hat{n}^{\hat{e}}} = \Theta_{\hat{e}}w_{\hat{e}}^k - (\Lambda_{\alpha^{\hat{e}}}^{\hat{e}}w_{E^{\hat{e}}}^k)_{\hat{n}^{\hat{e}}} = (\Lambda_\alpha w^k)_{\hat{n}^{\hat{e}}},$$

where in the first equality we used the definition of $\tilde{w}^k$ and in the second equality the fact that $0 = \Theta_{\hat{e}}(\tilde{w}_{\hat{e}}^k - w_{\hat{e}}^k) - (\Lambda_{\alpha^{\hat{e}}}^{\hat{e}}z_{E^{\hat{e}}}^k)_{\hat{n}^{\hat{e}}}$. Also, since $0 = (\Lambda_{\alpha^{\hat{e}}}^{\hat{e}}z_{E^{\hat{e}}}^k)_n$ for all $n \in V^{\hat{e}} \setminus \{\hat{n}^{\hat{e}}\}$ we have

$$(\Lambda_\alpha \tilde{w}^k)_n = (\Lambda_{\alpha^{\hat{e}}}^{\hat{e}}(w_{E^{\hat{e}}}^k - z_{E^{\hat{e}}}^k))_n = (\Lambda_{\alpha^{\hat{e}}}^{\hat{e}}w_{E^{\hat{e}}}^k)_n = (\Lambda_\alpha w^k)_n$$

for $n \in V^{\hat{e}} \setminus \{\hat{n}^{\hat{e}}\}$, $\hat{e} \in \hat{E}$.

This implies that also $\Lambda_\alpha \tilde{w}^k \overset{*}{\rightharpoonup} y$ in X_V and, since $(\Theta_{\hat{E}}\tilde{w}_{\hat{E}}^k)_k$ is weakly* convergent, that also $(\Lambda_{\alpha^{\hat{e}}}^{\hat{e}}\tilde{w}_{E^{\hat{e}}}^k)_k$, with $\tilde{w}_{E^{\hat{e}}}^k := w_{E^{\hat{e}}}^k - z_{E^{\hat{e}}}^k$, is weakly* convergent for each $\hat{e} \in \hat{E}$. By induction hypothesis, there hence exist $w_{E^{\hat{e}}} \in \mathrm{dom}(\Lambda_{\alpha^{\hat{e}}}^{\hat{e}})$ such that $\mathsf{w^*}\text{-}\lim_k \Lambda_{\alpha^{\hat{e}}}^{\hat{e}}\tilde{w}_{E^{\hat{e}}}^k = \Lambda_{\alpha^{\hat{e}}}^{\hat{e}}w_{E^{\hat{e}}}$. Defining $w = (w_{\hat{E}}, (w_{E^{\hat{e}}})_{\hat{e} \in \hat{E}})$ we finally see that $\Lambda_\alpha w = y$, since, from $\Lambda_\alpha \tilde{w}^k \overset{*}{\rightharpoonup} y$ it follows that

$$(\Lambda_\alpha w)_{\hat{n}} = Kw_{\hat{E}} = \mathsf{w^*}\text{-}\lim_k K\tilde{w}_{\hat{E}}^k = y_{\hat{n}},$$

$$(\Lambda_\alpha w)_{n^{\hat{e}}} = \Theta_{\hat{e}}w_{\hat{e}} - (\Lambda_{\alpha^{\hat{e}}}^{\hat{e}}w_{E^{\hat{e}}}^k)_{\hat{n}^{\hat{e}}} = \mathsf{w^*}\text{-}\lim_k \left(\Theta_{\hat{e}}\tilde{w}_{\hat{e}}^k - (\Lambda_{\alpha^{\hat{e}}}^{\hat{e}}\tilde{w}_{E^{\hat{e}}}^k)_{n^{\hat{e}}}\right) = y_{n^{\hat{e}}}$$

for each $\hat{e} \in \hat{E}$, and

$$(\Lambda_\alpha w)_n = (\Lambda_{\alpha^{\hat{e}}}^{\hat{e}}w_{E^{\hat{e}}})_n = \mathsf{w^*}\text{-}\lim_k (\Lambda_{\alpha^{\hat{e}}}^{\hat{e}}\tilde{w}_{E^{\hat{e}}}^k)_n = y_n$$

for each $n \in V^{\hat{e}} \setminus \{n^{\hat{e}}\}$, $\hat{e} \in \hat{E}$. This completes the proof. $\square$

5.1 Examples of predual regularization graph functionals

Here we provide predual regularization graph functionals for several examples introduced in Section 2.1 by verifying the additional assumption (H9). We represent such predual regularization graphs as in Figure 5. In this context, we denote by $\mathcal{I}_{B_p}$ the indicator function of the L^p unit ball for $p \in [1, \infty]$. Note that, for sake of clarity, in the root node of each predual regularization graph we write the corresponding functional $v_{\hat{n}} \mapsto \Psi_{\hat{n}}^{\#}(v_{\hat{n}}) - \langle u, v_{\hat{n}} \rangle$

and not just $\Psi_{\hat{n}}^{\#}$. Moreover, nodes represented by an empty circle are associated with the zero functional.

Total variation. Figure 5a shows a predual regularization graph for TV. We refer to Section 2.1 for the construction of the regularization graph realizing TV. We remind also that $X_1 = L^p(\Omega)$ and $X_2 = \mathcal{M}(\Omega, \mathbb{R}^d)$ where $1 < p \leq d'$ with $d' = d/(d-1)$ in case $d > 1$ and $d' = \infty$ else. Therefore the predual Banach spaces associated with the nodes are $X_1^{\#} = L^{p'}(\Omega)$ and $X_1^{\#} = C_0(\Omega, \mathbb{R}^d)$ where p' satisfies $1/p + 1/p' = 1$. Moreover, it is easy to see that the convex conjugate of $\mathcal{I}_{B_\infty}$ on $C_0(\Omega, \mathbb{R}^d)$ is $\|\cdot\|_{\mathcal{M}}$, and the convex conjugate of the zero function is $\mathcal{I}_{\{0\}}$. To show (H9), we further claim that $\nabla : \mathrm{BV}(\Omega) \subset L^{d'}(\Omega) \to \mathcal{M}(\Omega, \mathbb{R}^d)$ is the adjoint of the negative divergence operator $-\mathrm{div} : \mathrm{dom}(-\mathrm{div}) \subset C_0(\Omega, \mathbb{R}^d) \to L^d(\Omega)$ defined in the weak sense as

$$\langle -\mathrm{div}\,\varphi, \psi \rangle = \int_\Omega \varphi \cdot \nabla \psi \, \mathrm{d}x \quad \forall \psi \in C^1(\Omega) \tag{5.9}$$

with domain

$$\mathrm{dom}(-\mathrm{div}) = \{\varphi \in C_0(\Omega, \mathbb{R}^d) \,|\, \mathrm{div}\,\varphi \in L^d(\Omega)\}. \tag{5.10}$$

Note that from (5.9), using a simple density argument we obtain that $\mathrm{dom}(-\mathrm{div})$ is densely defined and closed in $C_0(\Omega, \mathbb{R}^d)$. To show that ∇ is the adjoint of $-\mathrm{div}$ according to the definitions above, it is enough to observe that

$$\mathrm{dom}(-\mathrm{div}^*) = \left\{ u \in L^{d'}(\Omega) \,\middle|\, \exists c > 0 \text{ such that } \int_\Omega u\,\mathrm{div}\,\varphi\,\mathrm{d}x \leq c\|\varphi\|_\infty \,\forall \varphi \in \mathrm{dom}(-\mathrm{div}) \right\}$$

implying that $\mathrm{dom}(-\mathrm{div}^*) = \mathrm{BV}(\Omega)$ and $(-\mathrm{div})^* = \nabla$. The predual regularization graph functional $R_\alpha^{\#} : L^p(\Omega) \to [0, +\infty]$ is given as

$$R_\alpha^{\#}(u) = -\inf \left\{ -\langle u, v_1 \rangle + \mathcal{I}_{B_\infty}(v_2) \,\middle|\, v_1 \in L^{p'}(\Omega),\, v_2 \in \mathrm{dom}(-\mathrm{div}),\, v_1 = -\mathrm{div}\,v_2 \right\}$$

$$= \sup \left\{ \int_\Omega u\,\mathrm{div}\,v\,\mathrm{d}x \,\middle|\, v \in C_0(\Omega, \mathbb{R}^d),\, \mathrm{div}\,v \in L^{p'}(\Omega),\, \|v\|_\infty \leq 1 \right\}.$$

Infimal convolution of $\mathrm{TV}^{k_1} - \mathrm{TV}^{k_2}$. Figure 5b shows a predual regularization graph corresponding to the infimal convolution of TV^{k_1} and TV^{k_2} with $k_1, k_2 \in \mathbb{N}$. We refer to Section 2.1 for the construction of the regularization graph realizing the infimal convolution of $\mathrm{TV}^{k_1} - \mathrm{TV}^{k_2}$. We remind also that p is chosen such that $1 < p \leq \min\{d_1', d_2'\}$, where $d_i' = d/(d-k_i)$ in case $d_i > k_i$ and $d_i' = \infty$ else. Similarly to the TV predual regularization graph, the pre-adjoint of each linear operator $\nabla^{k_i} : L^{d_i'}(\Omega) \to \mathcal{M}(\Omega, \mathrm{Sym}^{k_i}(\mathbb{R}^d))$ can be seen to be the (possibly negative) higher-order divergence $(-1)^{k_i}\mathrm{div}^{k_i} : \mathrm{dom}((-1)^{k_i}\mathrm{div}^{k_i}) \subset C_0(\Omega, \mathrm{Sym}^{k_i}(\mathbb{R}^d)) \to L^{d_i}(\Omega)$ defined in the weak sense as

$$\langle (-1)^{k_i}\mathrm{div}^{k_i}\varphi, \psi \rangle = (-1)^{k_i+1} \int_\Omega \varphi \cdot \nabla^{k_i}\psi \, \mathrm{d}x \quad \forall \psi \in C^{k_i}(\Omega) \tag{5.11}$$

with domain

$$\mathrm{dom}((-1)^{k_i}\mathrm{div}^{k_i}) = \{\varphi \in C_0(\Omega, \mathrm{Sym}^{k_i}(\mathbb{R}^d)) \,|\, \mathrm{div}^{k_i}\varphi \in L^{d_i}(\Omega)\} \tag{5.12}$$

which is again closed and densely defined, showing (H9). The predual regularization graph functional $R_\alpha^\# : L^p(\Omega) \to [0, +\infty]$ is given as

$$R_\alpha^\#(u) = -\inf \left\{ -\langle u, v_1 \rangle + \mathcal{I}_{B_\infty}(v_2) + \mathcal{I}_{B_\infty}(v_3) \,\middle|\, v_1 \in L^{p'}(\Omega), \ v_2 \in \mathrm{dom}((-1)^{k_1}\mathrm{div}^{k_1}), \right.$$
$$\left. v_3 \in \mathrm{dom}((-1)^{k_2}\mathrm{div}^{k_2}), \ v_1 = (-1)^{k_1}\mathrm{div}^{k_1} v_2, \ \alpha v_1 = (-1)^{k_2}\mathrm{div}^{k_2} v_3 \right\}$$
$$= \sup \left\{ \int_\Omega u\,\mathrm{div}^{k_1} u_1 \,\mathrm{d}x \,\middle|\, u_i \in C_0(\Omega, \mathrm{Sym}^{k_i}(\mathbb{R}^d)), \ \mathrm{div}^{k_i} u_i \in L^{p'}(\Omega) \right.$$
$$\left. \mathrm{div}^{k_2} u_2 = \alpha\,\mathrm{div}^{k_1} u_1, \ \|u_i\|_\infty \leq 1, \ i = 1, 2 \right\}.$$

Total generalized variation of order 2. Figure 5c shows a predual regularization graph for TGV_α^2. We refer to Section 2.1 for the construction of the regularization graph realizing the total generalized variation of order 2. We remind also that p is chosen such that $1 < p \leq d'$ where $d' = d/(d-1)$ in case $d > 1$ and $d' = \infty$ else. The pre-adjoint of $\nabla : \mathrm{BV}(\Omega) \subset L^{d'}(\Omega) \to \mathcal{M}(\Omega, \mathbb{R}^d)$ is given as in the TV example. Moreover, a pre-adjoint of $\mathcal{E} : \mathrm{BD}(\Omega) \subset L^{d'}(\Omega, \mathbb{R}^d) \to \mathcal{M}(\Omega, \mathrm{Sym}^2(\mathbb{R}^d))$ is the negative divergence operator $-\mathrm{div} : \mathrm{dom}(-\mathrm{div}) \subset C_0(\Omega, \mathrm{Sym}^2(\mathbb{R}^d)) \to L^d(\Omega, \mathbb{R}^d)$ defined in the weak sense as in (5.9) with domain

$$\mathrm{dom}(-\mathrm{div}) = \{\varphi \in C_0(\Omega, \mathrm{Sym}^2(\mathbb{R}^d)) \,|\, \mathrm{div}\,\varphi \in L^d(\Omega, \mathbb{R}^d)\}, \tag{5.13}$$

which is again densely defined and closed, showing (H9). The predual regularization graph functional $R_\alpha^\# : L^p(\Omega) \to [0, +\infty]$ is for $\alpha > 0$ given as

$$R_\alpha^\#(u) = -\inf \left\{ -\langle u, v_1 \rangle + \mathcal{I}_{\{\|v\|_\infty \leq 1\}}(v_3) + \mathcal{I}_{\{\|v\|_\infty \leq 1\}}(v_4) \,\middle|\, v_1 \in L^{p'}(\Omega), \ v_2 \in \mathrm{dom}(-\mathrm{div}), \right.$$
$$\left. v_3 \in C_0(\Omega, \mathbb{R}^d), \ v_4 \in \mathrm{dom}(-\mathrm{div}), \ v_1 = -\mathrm{div}\,v_2, \ v_2 = v_3, \ \alpha v_2 = -\mathrm{div}\,v_4 \right\}$$
$$= \sup \left\{ \int_\Omega u\,\mathrm{div}^2 v \,\mathrm{d}x \,\middle|\, v \in C_0(\Omega, \mathrm{Sym}^2(\mathbb{R}^d)), \ \mathrm{div}\,v \in C_0(\Omega, \mathbb{R}^d), \ \mathrm{div}^2 v \in L^{p'}(\Omega, \mathbb{R}^d) \right.$$
$$\left. \|\mathrm{div}\,v\|_\infty \leq 1, \ \|\alpha v\|_\infty \leq 1 \right\}.$$

TGV^2**-shearlet infimal convolution.** Figure 5d shows a predual regularization graph for the TGV^2-shearlet infimal convolution model. We refer to Section 2.1 for the construction of the regularization graph realizing for the TGV^2-shearlet infimal convolution. We also remind that the exponent p is chosen as $1 < p \leq 2$. Note that a predual of the extension to infinity of the ℓ^1 norm is the indicator function of the unit ball of c_0, denoted by $\mathcal{I}_{c_0} : \ell^2(\mathbb{Z}^4) \to [0, +\infty]$. Thanks to the closedness of $c_0 \cap \ell^2$ in ℓ^2, such indicator function is lower semicontinuous. Moreover, as $\mathcal{SH} : L^2(\mathbb{R}^2) \to \ell^2(\mathbb{Z}^4)$ defined according to (2.5) is a bounded operator between Hilbert spaces, its pre-adjoint exists and is bounded, showing (H9). Further, it can be easily characterized for $v \in \ell^2(\mathbb{Z}^4)$ as

$$\mathcal{SH}^\# v = \sum_{j,k \in \mathbb{Z}, \, m \in \mathbb{Z}^2} v_{j,k,m} \Psi_{j,k,m}.$$

Finally, noticing the the pre-adjoint of the restriction operator r_Ω is the extension to zero outside Ω (denoted by r_Ω^0), the predual regularization graph functional $R_\alpha^\# : L^p(\Omega) \to [0, +\infty]$

is for $\alpha_0, \alpha_1 > 0$ given as

$$
\begin{aligned}
R_\alpha^\#(u) = -\inf\Big\{ &-\langle u, v_1\rangle + \mathcal{I}_{\{\|v\|_\infty \leq 1\}}(v_3) + \mathcal{I}_{\{\|v\|_\infty \leq 1\}}(v_4) + \mathcal{I}_{c_0}(v_5)\,\big|\, v_1 \in L^{p'}(\Omega), \\
&v_2 \in \mathrm{dom}(-\mathrm{div}),\ v_3 \in C_0(\Omega, \mathbb{R}^2),\ v_4 \in \mathrm{dom}(-\mathrm{div}),\ v_5 \in \ell^2(\mathbb{Z}^4),\ v_1 = -\mathrm{div}\,v_2, \\
&\alpha_0 r_\Omega^0 v_1 = \mathcal{SH}^\# v_5,\ v_2 = v_3,\ \alpha_1 v_2 = -\mathrm{div}\,v_4\Big\} \\
= \sup\Big\{ &\int_\Omega u\,\mathrm{div}^2 v\,\mathrm{d}x\,\Big|\, v \in C_0(\Omega, \mathrm{Sym}(\mathbb{R}^2)),\ \mathrm{div}\,v \in C_0(\Omega, \mathbb{R}^2),\ \mathrm{div}^2 v \in L^{p'}(\Omega, \mathbb{R}^2), \\
&z \in c_0(\mathbb{Z}^4),\ \|\mathrm{div}\,v\|_\infty \leq 1,\ \|\alpha_1 v\|_\infty \leq 1,\ \alpha_0 r_\Omega^0 \mathrm{div}^2 v = \mathcal{SH}^\# z,\ \|z\|_\infty \leq 1\Big\}.
\end{aligned}
$$

6 Regularization of linear inverse problems

6.1 Setting and well-posedness

We now consider the application of regularization graphs to the regularization of linear inverse problems. That is, with $K : X_{\hat{n}} \to Y$ a bounded linear operator (the forward model), $S_f : Y \to [0, \infty)$ a discrepancy functional associated with the data f and $\beta > 0$ a regularization parameter, we consider the minimization problem

$$
\min_{u \in X_{\hat{n}}} S_f(Ku) + \beta R_\alpha(u), \tag{6.1}
$$

where $R_\alpha = R(\mathcal{G}_\alpha) : X_{\hat{n}} \to [0, \infty]$ with $\mathcal{G}_\alpha$ a regularization graph with root node $\hat{n}$.

Remark 6.1 (Forward operator with general domain X). Note that considering only forward operators being defined on $X_{\hat{n}}$, where bounded sequences need to admit weak* convergent subsequences according to (H8), is not a restriction compared to considering general operators $\tilde{K} : X \to Y$ with X a Banach space such that $X_{\hat{n}} \hookrightarrow X$ and R_α being extended by ∞ to X as in Proposition 4.8, since one can always recover this setting by choosing $K = \tilde{K} \circ I_{X_{\hat{n}}, X}$, with $I_{X_{\hat{n}}, X}$ the continuous embedding of $X_{\hat{n}}$ to X.

In order to study convergence in the data space for general discrepancies S_f, we introduce the following notion of convergence: We say the functionals $(S_{f^k})_k$ converge to S_f if
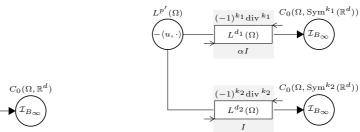
$$
\begin{cases}
S_f(v) \leq \liminf_k S_{f^k}(v^k) & \text{whenever } v^k \rightharpoonup v \text{ in } Y, \\
S_f(v) \geq \limsup_k S_{f^k}(v) & \text{for each } v \in Y.
\end{cases} \tag{6.2}
$$
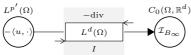
Further, we say that $(S_{f^k})_k$ is *equi-coercive* if there exists a coercive function $S_0 : Y \to [0, \infty)$ such that $S_{f^k} \geq S_0$ in Y for all k. Note that this always holds true if $S_{f^k}(v) := \frac{1}{q}\|v - f^k\|_Y^q$ and $f^k \to f$ in Y, but the more general assumptions allow us to capture, for instance, also the situation when S_f is the Kullback–Leibler divergence [7, Example 2.16].

Now, under weak assumptions, the previously established properties of R_α allow to obtain a standard well-posedness result for (6.1).
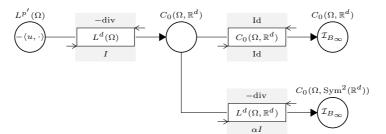
Theorem 6.2. *Let $R_\alpha = R(\mathcal{G}_\alpha)$ with $\mathcal{G}_\alpha$ being a regularization graph with weights α and root node $\hat{n}$ such that $X_{\hat{n}}$ is reflexive, $\beta > 0$, and let Y be a Banach space, $K : X_{\hat{n}} \to Y$ be linear and continuous and $S_f : Y \to [0, \infty]$ be a proper, convex, weakly lower semi-continuous and*
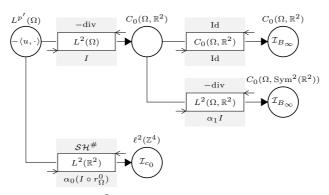
(a) Total variation (TV).

(b) $TV^{k_1} - TV^{k_2}$ infimal convolution.

(c) Second order total generalized variation.

(d) TGV^2-shearlet infimal convolution.

Figure 5: Examples of predual regularization graphs.

33

coercive discrepancy functional. Then, the Tikhonov minimization problem (6.1) *is well-posed, i.e., there exists a solution and the solution mapping is stable in sense that, if S_{f^k} converges to S_f and $(S_{f^k})_k$ is equi-coercive, then for each sequence of minimizers $(u^k)_k$ of* (6.1) *with discrepancy S_{f^k},*

   i) either $S_{f^k}(Ku^k) + \beta R_\alpha(u^k) \to \infty$ as $k \to \infty$ and (6.1) *with discrepancy S_f does not admit a finite solution,*

   ii) or $S_{f^k}(Ku^k) + \beta R_\alpha(u^k) \to \min_{u \in X_{\hat{n}}} S_f(Ku) + \beta R_\alpha(u)$ as $k \to \infty$ and there exists, possibly up to shifts by functions in $\ker(K) \cap L$, with L the invariant subspace of R_α, a weak accumulation point $u \in X_{\hat{n}}$ of $(u^k)_k$ that minimizes (6.1) *with discrepancy S_f.*

Further, in case (6.1) *with discrepancy S_f admits a finite solution, for each subsequence $(u^{k_i})_i$ weakly converging to some $u \in X_{\hat{n}}$, it holds that $R_\alpha(u^{k_i}) \to R_\alpha(u)$ as $i \to \infty$. Also, if S_f is strictly convex and K is injective, finite solutions u of* (6.1) *are unique and $u^k \rightharpoonup u$ in $X_{\hat{n}}$.*

Proof. Existence follows by the application of the direct method of calculus of variations in $X_{\hat{n}}$. More precisely, given a minimizing sequence $(u^k)_k$ for (6.1) we can apply Lemma 4.2 with $W = X = X_{\hat{n}}$, $F = R_\alpha$, $\Theta = \text{Id}$ and L being the finite dimensional invariant space of R_α provided by Theorem 4.3, to obtain the existence of another minimizing sequence $(\tilde{u}^k)_k$ for (6.1) that is bounded in $X_{\hat{n}}$. Note that the assumptions of Lemma 4.2 are fulfilled since the weak lower semi-continuity of R_α (which is equivalent to weak* lower semi-continuity of R_α by reflexivity of $X_{\hat{n}}$) and Assumption i) of Lemma 4.2 hold as a consequence of Theorem 4.3, and the boundedness of $(Ku^k)_k$ follows from the coercivity of S_f. Therefore, thanks to the weak lower semi-continuity of R_α and the boundedness of K we can apply the direct method to the sequence $(\tilde{u}^k)_k$ and conclude existence of minimizers for (6.1).

The claimed stability follows with standard arguments. For instance, it can be proven adapting straightforwardly [7, Theorem 2.14]. □

Remark 6.3. Note that the results of Theorem 6.2 can also be modified to hold without assuming reflexivity of $X_{\hat{n}}$ but assuming, for instance, that K is weak*-to-weak continuous. Indeed, in this setting, Lemma 4.2 applies the same way and existence follows from the coercivity statement of Lemma 4.2 using weak*-to-weak continuity of K and weak lower semi-continuity of S_f. Likewise, also the claimed stability can be shown by straightfoward adaptions.

6.2 Convergence and stability for varying parameters

In this section we study the stability of solutions of (6.1) for varying parameters α and for vanishing noise. To this aim, we first define a variant of regularization graphs for vanishing weights.

Definition 6.4. *For $\mathcal{G}_\alpha$ a regularization graph with underlying graph $G = (V, E)$ and $\alpha \in [0, \infty)^E$ weights, define $\tilde{\alpha} = (\tilde{\alpha}_e)_{e \in E}$ as $\tilde{\alpha}_e = \alpha_e$ for $\alpha_e > 0$ and $\tilde{\alpha}_e = 1$ else. Further, for $n \in V$, set $\tilde{\Psi}_n = \mathcal{I}_{\{0\}}$ in case that $n \neq \hat{n}$ and $\alpha_{(n^-, n)} = 0$, and $\tilde{\Psi}_n = \Psi_n$ else. Then, we define $\hat{\mathcal{G}}_\alpha = \tilde{\mathcal{G}}_{\tilde{\alpha}}$ with $\tilde{\mathcal{G}} = (G, (\tilde{\Psi}_n)_{n \in V}, (\Theta_e)_{e \in E}, (\Phi_e)_{e \in E})$, that is, $\hat{\mathcal{G}}_\alpha$ is the regularization graph $\mathcal{G}_\alpha$ with zero weights being replaced by 1 and node functionals associated with tails of zero-weight edges being replaced by $\mathcal{I}_{\{0\}}$. Finally, we set $\hat{R}_\alpha = R(\hat{\mathcal{G}}_\alpha)$, being the regularization graph functional associated to $\hat{\mathcal{G}}_\alpha$.*

This definition is required to deal with lower semi-continuity with respect to weights $(\alpha_e)_e$ converging to zero. An example of $\hat{R}_\alpha$ in case $R_\alpha(u) = R(\mathcal{G}_\alpha)(u) = \min_{w \in \mathrm{BD}(\Omega, \mathbb{R}^d)} \|\nabla u - \alpha_0 w\|_\mathcal{M} + \|\mathcal{E}w\|_\mathcal{M}$ and $\alpha_0 = 0$ is given as

$$\hat{R}_\alpha(u) = \min_{w \in \mathrm{BD}(\Omega, \mathbb{R}^d)} \|\nabla u - w\|_\mathcal{M} \quad \text{s.t. } \mathcal{E}w = 0.$$

It is easy to see that for any regularization graph $\mathcal{G}_\alpha$ and any choice of weights α, $\hat{\mathcal{G}}_\alpha$ is again a regularization graph such that all previous results apply. Moreover, the following lemma holds.

Lemma 6.5. *Let $\mathcal{G}_\alpha$ be a regularization graph with root node $\hat{n}$ and weights $\alpha \in [0, \infty)^E$, $R_\alpha = R(\mathcal{G}_\alpha)$ and $\hat{R}_\alpha = R(\hat{\mathcal{G}}_\alpha)$. Then for every $u \in X_{\hat{n}}$ we have that $\hat{R}_\alpha(u) \le R_\alpha(u)$.*

Proof. Arguing by induction we suppose that the claimed assertion holds for any regularization graph of height less than h and we assume that the height of $\mathcal{G}_\alpha$ is h. For $h = 0$, the result holds since $\hat{R}_\alpha = R_\alpha$, so we assume $h \ge 1$. Using the recursive representation of R_α and the notation from Lemma 3.2 it holds that

$$R_\alpha(u) = \inf \left\{ \Psi_{\hat{n}} \left(u - \sum_{\hat{e} \in \hat{E}: \alpha_{\hat{e}} > 0} \alpha_{\hat{e}} \Phi_{\hat{e}} w_{\hat{e}} \right) + \sum_{\hat{e} \in \hat{E}} R^{\hat{e}}_{\alpha^{\hat{e}}}(\Theta_{\hat{e}} w_{\hat{e}}) \,\middle|\, w_{\hat{e}} \in \mathrm{dom}(\Theta_{\hat{e}}) \text{ for all } \hat{e} \in \hat{E} \right\}$$
$$(6.3)$$

where, for $\hat{e} \in \hat{E}$, $R^{\hat{e}}_{\alpha^{\hat{e}}} = R(\mathcal{G}^{\hat{e}}_{\alpha^{\hat{e}}})$ and $\mathcal{G}^{\hat{e}}_{\alpha^{\hat{e}}} = (\mathcal{G}^{\hat{e}}, \alpha^{\hat{e}})$, with $\mathcal{G}^{\hat{e}} = (G^{\hat{e}}, (\Psi_n)_{n \in V^{\hat{e}}}, (\Theta_e)_{e \in E^{\hat{e}}}, (\Phi_e)_{e \in E^{\hat{e}}})$ and $\alpha^{\hat{e}} = (\alpha_e)_{e \in E^{\hat{e}}}$, is a regularization graph of height at most $h-1$ with associated graph $G^{\hat{e}} = (V^{\hat{e}}, E^{\hat{e}})$ that is a subtree of $G = (V, E)$ with $\hat{n}^{\hat{e}}$ as root node. Similarly, using Lemma 3.2, we write $\hat{R}_\alpha(u)$ as

$$\hat{R}_\alpha(u) = \inf \left\{ \Psi_{\hat{n}} \left(u - \sum_{\hat{e} \in \hat{E}: \alpha_{\hat{e}} > 0} \alpha_{\hat{e}} \Phi_{\hat{e}} w_{\hat{e}} - \sum_{\hat{e} \in \hat{E}: \alpha_{\hat{e}} = 0} \Phi_{\hat{e}} w_{\hat{e}} \right) + \sum_{\hat{e} \in \hat{E}: \alpha_{\hat{e}} = 0} \bar{R}^{\hat{e}}_{\tilde{\alpha}^{\hat{e}}}(\Theta_{\hat{e}} w_{\hat{e}}) \right.$$
$$\left. + \sum_{\hat{e} \in \hat{E}: \alpha_{\hat{e}} > 0} \bar{R}^{\hat{e}}_{\tilde{\alpha}^{\hat{e}}}(\Theta_{\hat{e}} w_{\hat{e}}) \,\middle|\, w_{\hat{e}} \in \mathrm{dom}(\Theta_{\hat{e}}) \text{ for all } \hat{e} \in \hat{E} \right\}, \qquad (6.4)$$

where, for $\hat{e} \in \hat{E}$, $\bar{R}^{\hat{e}}_{\tilde{\alpha}^{\hat{e}}} = R(\bar{\mathcal{G}}^{\hat{e}}_{\tilde{\alpha}^{\hat{e}}})$ and $\bar{\mathcal{G}}^{\hat{e}}_{\tilde{\alpha}^{\hat{e}}} = (\bar{\mathcal{G}}^{\hat{e}}, \tilde{\alpha}^{\hat{e}})$, with $\bar{\mathcal{G}}^{\hat{e}} = (G^{\hat{e}}, (\tilde{\Psi}_n)_{n \in V^{\hat{e}}}, (\Theta_e)_{e \in E^{\hat{e}}}, (\Phi_e)_{e \in E^{\hat{e}}})$ and $(\tilde{\alpha}_e)_{e \in E^{\hat{e}}}$ according to Definition 6.4, is a regularization graph of height at most $h - 1$ with associated graphs $G^{\hat{e}} = (V^{\hat{e}}, E^{\hat{e}})$. Note that $\tilde{\Psi}_{\hat{n}^{\hat{e}}} = \Psi_{\hat{n}^{\hat{e}}}$ for every $\hat{e} \in \hat{E}$ such that $\alpha_{\hat{e}} > 0$. Therefore, by the way the modified regularization graph $\hat{\mathcal{G}}^{\hat{e}}_{\alpha^{\hat{e}}}$ is obtained from $\mathcal{G}^{\hat{e}}_{\alpha^{\hat{e}}}$ according to Definition 6.4, it follows that, for $\hat{e} \in \hat{E}$ with $\alpha_{\hat{e}} > 0$, $\bar{R}^{\hat{e}}_{\tilde{\alpha}^{\hat{e}}} = \hat{R}^{\hat{e}}_{\alpha^{\hat{e}}}$ with $\hat{R}^{\hat{e}}_{\alpha^{\hat{e}}} = R(\hat{\mathcal{G}}^{\hat{e}}_{\alpha^{\hat{e}}})$.

Thus, choosing $w_{\hat{e}} = 0$ for every $\hat{e} \in \hat{E}$ such that $\alpha_{\hat{e}} = 0$ in (6.4) and using the induction hypothesis as well as (6.3), we estimate for every $u \in X_{\hat{n}}$ as follows

$$\hat{R}_\alpha(u) \le \inf \left\{ \Psi_{\hat{n}} \left(u - \sum_{\hat{e} \in \hat{E}: \alpha_{\hat{e}} > 0} \alpha_{\hat{e}} \Phi_{\hat{e}} w_{\hat{e}} \right) + \sum_{\hat{e} \in \hat{E}: \alpha_{\hat{e}} > 0} \hat{R}^{\hat{e}}_{\alpha^{\hat{e}}}(\Theta_{\hat{e}} w_{\hat{e}}) \,\middle|\, w_{\hat{e}} \in \mathrm{dom}(\Theta_{\hat{e}}), \right.$$
$$\left. \forall \hat{e} \in \hat{E} \text{ such that } \alpha_{\hat{e}} > 0 \right\} \le R_\alpha(u).$$

$\square$

We now prove a weak* lower semi-continuity result for regularization graph functionals with respect to the parameters.

Theorem 6.6. *Let $\mathcal{G}_\alpha$ be a regularization graph with root node $\hat{n}$ and weights $\alpha \in [0, \infty)^E$, $R_\alpha = R(\mathcal{G}_\alpha)$, and $(\alpha^k)_k$ be a sequence of weights in $(0, \infty)^E$ such that $(\alpha^k)_k \to \alpha$. Then, for every sequence $(u^k)_k$ in $X_{\hat{n}}$ such that $u^k \overset{*}{\rightharpoonup} u \in X_{\hat{n}}$ it holds that*

$$\hat{R}_\alpha(u) \leq \liminf_{k \to \infty} R_{\alpha^k}(u^k). \tag{6.5}$$

Moreover, for $u \in X_{\hat{n}}$ and

$$\gamma_k := \min \Big\{ \prod_{e \in F} \alpha_e^k / \alpha_e \, \Big| \, F \subset E \text{ is either empty or} \tag{6.6}$$

$$\text{a chain with root } \hat{n} \in V \text{ and } \alpha_e > 0 \; \forall e \in F \Big\}, \tag{6.7}$$

using again the conventions that for $F = \emptyset$, $\prod_{e \in \emptyset} \frac{\alpha_e^2}{\alpha_e^1} = 1$, we have that $\gamma_k \leq 1$, $\gamma_k \to 1$,

$$R_{\alpha^k}(\gamma_k u) \leq \hat{R}_\alpha(u) \text{ for all } k \in \mathbb{N} \quad \text{and} \quad \lim_{k \to \infty} R_{\alpha^k}(\gamma_k u) = \hat{R}_\alpha(u). \tag{6.8}$$

Remark 6.7. Note that, in case each node functional Ψ_n is positively one homogeneous (such that R_{α^k} is positively one homogeneous according to Proposition 2.6), the convergence of (6.8) implies that $\lim_{k \to \infty} R_{\alpha^k}(u) = \hat{R}_\alpha(u)$. Also, in case $\alpha_e > 0$ for each $e \in E$, $\hat{R}_\alpha$ can be replaced by R_α.

Proof of Theorem 6.6. We argue again by induction and, supposing that the claimed assertions hold for any regularization graph of height less than h, assume that the height of $\mathcal{G}_\alpha$ is h. Again, for $h = 0$, the result holds trivially, so we assume $h \geq 1$.

We first deal with lower semi-continuity of $R_\alpha = R(\mathcal{G}_\alpha)$, for which, up to taking a non-relabeled subsequence, we assume that $\liminf_{k \to \infty} R_{\alpha^k}(u^k) = \lim_{k \to \infty} R_{\alpha^k}(u^k) < +\infty$. Using the recursive representation of R_{α^k} and the notation from Lemma 3.2, and the result of Theorem 4.3 we can select a sequence $(w^k)_k$ in $\bigtimes_{\hat{e} \in \hat{E}} \mathrm{dom}(\Theta_{\hat{e}})$ such that

$$R_{\alpha^k}(u^k) = \Psi_{\hat{n}} \left(u^k - \sum_{\hat{e} \in \hat{E}} \alpha_{\hat{e}}^k \Phi_{\hat{e}} w_{\hat{e}}^k \right) + \sum_{\hat{e} \in \hat{E}} R_{(\alpha^k)^{\hat{e}}}^{\hat{e}}(\Theta_{\hat{e}} w_{\hat{e}}^k), \tag{6.9}$$

with $R_{(\alpha^k)^{\hat{e}}}^{\hat{e}} = R(\mathcal{G}_{(\alpha^k)^{\hat{e}}}^{\hat{e}})$ and $\mathcal{G}_{(\alpha^k)^{\hat{e}}}^{\hat{e}}$ being regularization graphs of height at most $h - 1$ with graph structure $G^{\hat{e}} = (V^{\hat{e}}, E^{\hat{e}})$ and root node $\hat{n}^{\hat{e}}$. Thanks to Proposition 4.7 and the weights α_e^k being positive, the invariant subspace $L^{\hat{e}}$ of $R_{(\alpha^k)^{\hat{e}}}^{\hat{e}}$ does not depend on k and for

$$C_{\hat{e}, \alpha^k} := \max \Big\{ \prod_{e \in F} \alpha_e^k \, \big| \, F \subset E^{\hat{e}} \text{ is either empty or a chain with root } \hat{n}^{\hat{e}} \Big\}, \tag{6.10}$$

it follows that $(C_{\hat{e}, \alpha^k})_k$ is bounded and that

$$\|u - P_{\mathrm{rg}(\Theta_{\hat{e}}) \cap L^{\hat{e}}} u\|_{X_{\hat{n}^{\hat{e}}}} \leq C_{\hat{e}, \alpha^k} C_{\hat{e}} R_{(\alpha^k)^{\hat{e}}}^{\hat{e}}(u) + D_{\hat{e}} \quad \forall u \in X_{\hat{n}^{\hat{e}}} \tag{6.11}$$

with $C_{\hat{e}}, D_{\hat{e}}$ independent of k. Further, remember that each $\Theta_{\hat{e}}$ satisfies

$$\|w - P_{\ker(\Theta_{\hat{e}})}w\|_{X_{\hat{e}}} \leq B_{\hat{e}}\|\Theta_{\hat{e}}w\|_{X_{\hat{n}\hat{e}}} \quad \forall w \in \mathrm{dom}(\Theta_{\hat{e}}),$$

with $B_{\hat{e}} > 0$, thanks to Assumption (H6). Now defining $M^{\hat{e}} \subset \mathrm{dom}(\Theta_{\hat{e}})$ and $P^{\hat{e}} : \mathrm{dom}(\Theta_{\hat{e}}) \to M^{\hat{e}}$ as in (4.10) and (4.11), respectively, one sees that they also do not depend on k and, estimating as in (4.12), we obtain

$$\|w - P^{\hat{e}}w\|_{X_{\hat{e}}} \leq B_{\hat{e}}C_{\hat{e}}C_{\hat{e},\alpha^k}R^{\hat{e}}_{(\alpha^k)^{\hat{e}}}(\Theta_{\hat{e}}w) + B_{\hat{e}}D_{\hat{e}}. \tag{6.12}$$

Further, define $K : W \to X_{\hat{n}}$ with $W := \times_{\hat{e}\in\hat{E}} X_{\hat{e}}$ as

$$Kw = \sum_{\hat{e}\in\hat{E}} \Phi_{\hat{e}}w_{\hat{e}} \tag{6.13}$$

and $P_M w := (P^{\hat{e}}w_{\hat{e}})_{\hat{e}}$ for $w \in W$. Note that for $M = \times_{\hat{e}\in\hat{E}} M^{\hat{e}}$, $P_M : W \to M$ is a projection. Choose $P_{\ker(K)\cap M} : M \to \ker(K) \cap M$ to be a projection onto $\ker(K) \cap M$ and let $P_Z : M \to Z$ be defined as $P_Z = \mathrm{Id} - P_{\ker(K)\cap M}$, where $Z = \mathrm{rg}(\mathrm{Id} - P_{\ker(K)\cap M})$. Note that P_Z is a projection onto Z. Then, defining $\alpha_{\hat{E}}w := (\alpha_{\hat{e}}w_{\hat{e}})_{\hat{e}\in\hat{E}}$ for $w \in W$ and $\alpha_{\hat{E}} \in [0,\infty)^{\hat{E}}$, we can observe that, with $\alpha^k_{\hat{E}} = (\alpha^k_{\hat{e}})_{\hat{e}\in\hat{E}}$,

$$\tilde{w}^k := w^k - (P_M w^k - (1/\alpha^k_{\hat{E}})P_Z\alpha^k_{\hat{E}}P_M w^k) \tag{6.14}$$

also realizes the minimum in (6.9), since

$$w^k - \tilde{w}^k = P_M w^k - (1/\alpha^k_{\hat{E}})P_Z\alpha^k_{\hat{E}}P_M w^k = (1/\alpha^k_{\hat{E}})(\mathrm{Id} - P_Z)\alpha^k_{\hat{E}}P_M w^k$$

with $(\mathrm{Id} - P_Z)\alpha^k_{\hat{E}}P_M w^k \in \ker(K)\cap M$, such that $\sum_{\hat{e}\in\hat{E}} \alpha^k_{\hat{e}}\Phi_{\hat{e}}(w^k - \tilde{w}^k) = K(\mathrm{Id} - P_Z)\alpha^k_{\hat{E}}P_M w^k = 0$ and $\Theta_{\hat{e}}(w^k - \tilde{w}^k)_{\hat{e}} = (1/\alpha^k_{\hat{e}})\Theta_{\hat{e}}[(\mathrm{Id} - P_Z)\alpha^k_{\hat{E}}P_M w^k]_{\hat{e}} = 0$. By the estimate (6.12) we obtain for some constants $C, D, \tilde{D} > 0$ that

$$\|w^k - P_M w^k\|_W \leq C\left(\max_{\hat{e}\in\hat{E}} C_{\hat{e},\alpha^k}\right)\sum_{\hat{e}\in\hat{E}} R^{\hat{e}}_{(\alpha^k)^{\hat{e}}}(\Theta_{\hat{e}}w^k_{\hat{e}}) + D \leq \tilde{D} < \infty \tag{6.15}$$

where the constant $\tilde{D}$ does not depend on k thanks to the boundedness of $R_{\alpha^k}(u^k)$, the recursive formula (6.9), and the boundedness of $(C_{\hat{e},\alpha^k})_k$.

Since K is injective and bounded (see Remark 2.1) on the finite dimensional space Z, there exists $C > 0$ independent from k such that $\|z\|_W \leq C\|Kz\|_{X_{\hat{n}}}$ for all $z \in Z$. Thus we can estimate by coercivity of $\Psi_{\hat{n}}$ and using $P_Z\alpha^k_{\hat{E}}P_M w^k = \alpha^k_{\hat{E}}(\tilde{w}^k - (w^k - P_M w^k))$ that for generic constants $\tilde{C}, D$ (and $\tilde{D}$ as in (6.15)) we have

$$\begin{aligned}
\|P_Z\alpha^k_{\hat{E}}P_M w^k\|_W &\leq C\|KP_Z\alpha^k_{\hat{E}}P_M w^k\|_{X_{\hat{n}}} \\
&\leq C\|K\alpha^k_{\hat{E}}\tilde{w}^k\|_{X_{\hat{n}}} + \tilde{C}\|\alpha^k_{\hat{E}}w^k - \alpha^k_{\hat{E}}P_M w^k\|_W \\
&\leq C\|K\alpha^k_{\hat{E}}\tilde{w}^k - u^k\|_{X_{\hat{n}}} + C\|u^k\|_{X_{\hat{n}}} + \tilde{C}(\max_{\hat{e}\in\hat{E}}\alpha^k_{\hat{e}})\|w^k - P_M w^k\|_W \\
&\leq \tilde{C}\Psi_{\hat{n}}(u^k - K\alpha^k_{\hat{E}}\tilde{w}^k) + C\|u^k\|_{X_{\hat{n}}} + \tilde{C}(\max_{\hat{e}\in\hat{E}}\alpha^k_{\hat{e}})\tilde{D} \\
&\leq \tilde{C}R_{\alpha^k}(u^k) + D < \infty, \tag{6.16}
\end{aligned}$$

where also used (6.15), the fact that $(u^k)_k$ is uniformly bounded as it is weak* converging and that $\tilde{w}^k$ realizes the minimum in (6.9).

Now for $\hat{e} \in \hat{E}$ with $\alpha_{\hat{e}} > 0$ this together with (6.14) and (6.15) implies that $(\tilde{w}_{\hat{e}}^k)_k$ is bounded, hence admits a (non-relabeled) subsequence weak* converging to some $\tilde{w}_{\hat{e}} \in X_{\hat{e}}$ by (H8). Moreover, using (6.9), (6.11), (6.16), the boundedness of $R_{\alpha^k}(u^k)$ and the finite dimensionality of Z we have for $\hat{e} \in \hat{E}$ that

$$
\begin{aligned}
\|\Theta_{\hat{e}} \tilde{w}_{\hat{e}}^k\|_{X_{\hat{n}\hat{e}}} &\leq \|\Theta_{\hat{e}} w_{\hat{e}}^k - \Theta_{\hat{e}} P^{\hat{e}} w_{\hat{e}}^k\|_{X_{\hat{n}\hat{e}}} + \|\Theta_{\hat{e}}(1/\alpha_{\hat{e}}^k)(P_Z \alpha_{\hat{E}}^k P_M w^k)_{\hat{e}}\|_{X_{\hat{n}\hat{e}}} \\
&= \|\Theta_{\hat{e}} w_{\hat{e}}^k - P_{\mathrm{rg}(\Theta_{\hat{e}}) \cap L^{\hat{e}}} \Theta_{\hat{e}} w_{\hat{e}}^k\|_{X_{\hat{n}\hat{e}}} + \|\Theta_{\hat{e}}(1/\alpha_{\hat{e}}^k)(P_Z \alpha^k P_M w^k)_{\hat{e}}\|_{X_{\hat{n}\hat{e}}} \\
&\leq C_{\hat{e},\alpha^k} C_{\hat{e}} R_{(\alpha^k)^{\hat{e}}}^{\hat{e}}(\Theta_{\hat{e}} w_{\hat{e}}^k) + D_{\hat{e}} + C\|(P_Z \alpha^k P_M w^k)_{\hat{e}}\|_{X_{\hat{n}\hat{e}}} \\
&\leq \tilde{C} < +\infty,
\end{aligned}
$$

where the constant $\tilde{C}$ is independent of k and we use the definition of $P^{\hat{e}}$ in (4.11). Hence, by weak* sequential compactness of the $X_{\hat{n}\hat{e}}$ and weak*-closedness of $\Theta_{\hat{e}}$ we obtain $\tilde{w}_{\hat{e}} \in \mathrm{dom}(\Theta_{\hat{e}})$ and, up to taking a further non-relabeled subsequence, w*-$\lim_{k \to +\infty} \Theta \tilde{w}_{\hat{e}}^k = \Theta \tilde{w}_{\hat{e}}$.

Further, for $\hat{e} \in \hat{E}$ with $\alpha_{\hat{e}} = 0$, we see from (6.14), (6.15) and (6.16) that $(\alpha_{\hat{e}}^k \tilde{w}_{\hat{e}}^k)_k$ is bounded. Hence, up to taking a further subsequence, we can assume that

$$
\text{w*-}\lim_{k \to \infty} \alpha_{\hat{e}}^k \tilde{w}_{\hat{e}}^k = \text{w*-}\lim_{k \to \infty} (P_Z \alpha_{\hat{E}}^k P_M w^k)_{\hat{e}} = z_{\hat{e}} \in M^{\hat{e}} \tag{6.17}
$$

by (6.15) since $\alpha_{\hat{e}}^k \to 0$ as $k \to +\infty$. Using Lemma 3.2 we can write $\hat{R}_\alpha(u)$ as

$$
\begin{aligned}
\hat{R}_\alpha(u) = \inf \Bigg\{ &\Psi_{\hat{n}}\Bigg(u - \sum_{\hat{e} \in \hat{E}: \alpha_{\hat{e}} > 0} \alpha_{\hat{e}} \Phi_{\hat{e}} w_{\hat{e}} - \sum_{\hat{e} \in \hat{E}: \alpha_{\hat{e}} = 0} \Phi_{\hat{e}} w_{\hat{e}} \Bigg) + \sum_{\hat{e} \in \hat{E}: \alpha_{\hat{e}} = 0} \bar{R}_{\tilde{\alpha}^{\hat{e}}}^{\hat{e}}(\Theta_{\hat{e}} w_{\hat{e}}) \\
&+ \sum_{\hat{e} \in \hat{E}: \alpha_{\hat{e}} > 0} \bar{R}_{\tilde{\alpha}^{\hat{e}}}^{\hat{e}}(\Theta_{\hat{e}} w_{\hat{e}}) \Bigg| w_{\hat{e}} \in \mathrm{dom}(\Theta_{\hat{e}}) \text{ for all } \hat{e} \in \hat{E} \Bigg\},
\end{aligned}
$$

where, for $\hat{e} \in \hat{E}$, $\bar{R}_{\tilde{\alpha}^{\hat{e}}}^{\hat{e}} = R(\bar{\mathcal{G}}_{\tilde{\alpha}^{\hat{e}}}^{\hat{e}})$ and $\bar{\mathcal{G}}_{\tilde{\alpha}^{\hat{e}}}^{\hat{e}} = (\bar{\mathcal{G}}^{\hat{e}}, \tilde{\alpha}^{\hat{e}})$, with $\bar{\mathcal{G}}^{\hat{e}} = (G^{\hat{e}}, (\tilde{\Psi}_n)_{n \in V^{\hat{e}}}, (\Theta_e)_{e \in E^{\hat{e}}}, (\Phi_e)_{e \in E^{\hat{e}}})$ and $\tilde{\alpha}^{\hat{e}} = (\tilde{\alpha}_e)_{e \in E^{\hat{e}}}$ according to Definition 6.4, is a regularization graph of height at most $h - 1$ with graph structure $G^{\hat{e}} = (V^{\hat{e}}, E^{\hat{e}})$ and root node $\hat{n}^{\hat{e}}$. Note that for $\hat{e} \in \hat{E}$ such that $\alpha_{\hat{e}} > 0$ we have $\bar{R}_{\tilde{\alpha}^{\hat{e}}}^{\hat{e}} = \hat{R}_{\alpha^{\hat{e}}}^{\hat{e}}$ with $\hat{R}_{\alpha^{\hat{e}}}^{\hat{e}} = R(\hat{\mathcal{G}}_{\alpha^{\hat{e}}}^{\hat{e}})$ and $\hat{\mathcal{G}}_{\alpha^{\hat{e}}}^{\hat{e}}$ being the modification of the regularization graph $\mathcal{G}_{\alpha^{\hat{e}}}^{\hat{e}}$ according to Definition 6.4. Therefore, weak* lower semi-continuity of $\Psi_{\hat{n}}$, the induction hypothesis and (6.17), leading to $\hat{R}_{\alpha^{\hat{e}}}^{\hat{e}}(\Theta_{\hat{e}} z_{\hat{e}}) = 0$ for $\alpha_{\hat{e}} = 0$, then yields

$$
\begin{aligned}
\hat{R}_\alpha(u) &\leq \Psi_{\hat{n}}\Bigg(u - \sum_{\hat{e} \in \hat{E}: \alpha_{\hat{e}} > 0} \alpha_{\hat{e}} \Phi_{\hat{e}} \tilde{w}_{\hat{e}} - \sum_{\hat{e} \in \hat{E}: \alpha_{\hat{e}} = 0} \Phi_{\hat{e}} z_{\hat{e}} \Bigg) + \sum_{\hat{e} \in \hat{E}: \alpha_{\hat{e}} > 0} \hat{R}_{\alpha^{\hat{e}}}^{\hat{e}}(\Theta_{\hat{e}} \tilde{w}_{\hat{e}}) \\
&\leq \liminf_k \Psi_{\hat{n}}\Bigg(u^k - \sum_{\hat{e} \in \hat{E}} \alpha_{\hat{e}}^k \Phi_{\hat{e}} \tilde{w}_{\hat{e}}^k \Bigg) + \sum_{\hat{e} \in \hat{E}} R_{(\alpha^k)^{\hat{e}}}^{\hat{e}}(\Theta_{\hat{e}} \tilde{w}_{\hat{e}}^k) = \liminf_k R_{\alpha^k}(u^k).
\end{aligned}
$$

Now take $u \in X_{\hat{n}}$ and observe that, since the convergence $\gamma_k \to 1$ as $k \to +\infty$ is immediate, the second assertion of (6.8) follows directly from what we just showed, provided that $R_{\alpha^k}(\gamma_k u) \leq \hat{R}_\alpha(u)$ for every $k \in \mathbb{N}$ holds. In order to show the latter, we first select $w \in W$ to attain the minimum in the recursive representation of $\hat{R}_\alpha(u)$ according to Lemma 3.2 (which

38

is possible by Theorem 4.3), noting that we can choose $w_{\hat{e}} = 0$ for $\hat{e} \in \hat{E}$ with $\alpha_{\hat{e}} = 0$, and that $\alpha_{\hat{e}}^k w_{\hat{e}} \to \alpha_{\hat{e}} w_{\hat{e}}$ for all $\hat{e} \in \hat{E}$. In particular, $R^{\hat{e}}_{(\alpha^k)^{\hat{e}}}(\Theta_{\hat{e}} w_{\hat{e}}) = \hat{R}^{\hat{e}}_{\alpha_{\hat{e}}}(\Theta_{\hat{e}} w_{\hat{e}}) = 0$ for $\hat{e} \in \hat{E}$ with $\alpha_{\hat{e}} = 0$. Also, define

$$\gamma^{\hat{e}}_k := \min \Big\{ \prod_{e \in F} \alpha_e^k / \alpha_e \,\big|\, F \subset E^{\hat{e}} \text{ is either empty or a chain with root } \hat{n}^{\hat{e}} \text{ and } \alpha_e > 0 \ \forall e \in F \Big\},$$

using again the convention $\prod_{e \in \emptyset} \frac{\alpha_e^k}{\alpha_e} = 1$. Therefore, using the induction hypothesis together with Remark 2.1 and Proposition 2.6 we obtain

$$
\begin{aligned}
R_{\alpha^k}(\gamma_k u) &\leq \Psi_{\hat{n}} \left(\gamma_k u - \gamma_k \sum_{\hat{e} \in \hat{E}:\alpha_{\hat{e}}>0} \alpha_{\hat{e}}^k \Phi_{\hat{e}} \left(\frac{\alpha_{\hat{e}}}{\alpha_{\hat{e}}^k} w_{\hat{e}} \right) - \gamma_k \sum_{\hat{e} \in \hat{E}:\alpha_{\hat{e}}=0} \Phi_{\hat{e}} w_{\hat{e}} \right) \\
&\quad + \sum_{\hat{e} \in \hat{E}:\alpha_{\hat{e}}>0} R^{\hat{e}}_{(\alpha^k)^{\hat{e}}} \left(\Theta_{\hat{e}} \left(\gamma_k \frac{\alpha_{\hat{e}}}{\alpha_{\hat{e}}^k} w_{\hat{e}} \right) \right) \\
&\leq \gamma_k \Psi_{\hat{n}} \left(u - \sum_{\hat{e} \in \hat{E}:\alpha_{\hat{e}}>0} \alpha_{\hat{e}} \Phi_{\hat{e}}(w_{\hat{e}}) - \sum_{\hat{e} \in \hat{E}:\alpha_{\hat{e}}=0} \Phi_{\hat{e}} w_{\hat{e}} \right) \\
&\quad + \sum_{\hat{e} \in \hat{E}:\alpha_{\hat{e}}>0} \frac{\gamma_k}{\gamma_k^{\hat{e}}} \frac{\alpha_{\hat{e}}}{\alpha_{\hat{e}}^k} R^{\hat{e}}_{(\alpha^k)^{\hat{e}}}(\gamma_k^{\hat{e}} \Theta_{\hat{e}} w_{\hat{e}}) \\
&\leq \Psi_{\hat{n}} \left(u - \sum_{\hat{e} \in \hat{E}:\alpha_{\hat{e}}>0} \alpha_{\hat{e}} \Phi_{\hat{e}} w_{\hat{e}} - \sum_{\hat{e} \in \hat{E}:\alpha_{\hat{e}}=0} \Phi_{\hat{e}} w_{\hat{e}} \right) + \sum_{\hat{e} \in \hat{E}:\alpha_{\hat{e}}>0} \hat{R}^{\hat{e}}_{\alpha^{\hat{e}}}(\Theta_{\hat{e}} w_{\hat{e}}) = \hat{R}_{\alpha}(u),
\end{aligned}
$$

where we used that $\gamma_k \leq 1$ as well as $\frac{\gamma_k}{\gamma_k^{\hat{e}}} \frac{\alpha_{\hat{e}}}{\alpha_{\hat{e}}^k} \leq 1$ since $\gamma_k \leq \gamma_k^{\hat{e}} \frac{\alpha_{\hat{e}}^k}{\alpha_{\hat{e}}}$. $\qquad\square$

We are now ready to prove a result that will in particular imply stability for varying parameter α and convergence for vanishing noise for (6.1).

Theorem 6.8. *Let $R_\alpha = R(\mathcal{G}_\alpha)$ with $\mathcal{G}_\alpha$ be a regularization graph with weight α and root node $\hat{n}$ such that $X_{\hat{n}}$ is reflexive, and let Y be a Banach space, $K : X_{\hat{n}} \to Y$ be linear and continuous and $S_f, S_{f^k} : Y \to [0, \infty]$ for $k \in \mathbb{N}$ be proper, convex, lower semi-continuous and coercive discrepancy functionals with $S_f(v) = 0$ if and only if $v = f$. Further, assume that S_{f^k} converges to S_f and that $(S_{f^k})_k$ is equi-coercive. Choosing $\delta_k := S_{f^k}(f)$ (such that $\delta_k \to 0$ by convergence of S_{f^k}), let $(\alpha^k)_k$ in $(0,\infty)^E$ and $(\beta_k)_k$ in $(0,\infty)$ be such that*

i) $\beta_k \to \beta$, $\frac{\delta_k}{\beta_k} \to 0$, and $\quad \alpha_e^k \to \alpha_e$, $\alpha_e^k \geq \alpha_e$ for all $e \in E$.

In case $\beta = 0$, assume additionally that

ii) there exists $u_0 \in X_{\hat{n}}$ with $\hat{R}_\alpha(u_0) < +\infty$ such that $Ku_0 = f$.

Then, for $(u^k)_k$ a sequence of minimizers of (6.1) with parameters $(\beta_k)_k$ and $(\alpha^k)_k$, up to shifts in $\ker(K) \cap L$, with L being the invariant subspace of R_{α^k} (which does not depend on k), $(u^k)_k$ has a subsequence weakly converging in $X_{\hat{n}}$. Further, any limit $\hat{u}$ of a subsequence $(u^{k_i})_i$ converging weakly in $X_{\hat{n}}$ solves

$$\min_{u \in X_{\hat{n}}} S_f(Ku) + \beta \hat{R}_\alpha(u) \tag{6.18}$$

in case $\beta > 0$ and

$$\min_{u \in X_{\hat{n}}} \hat{R}_{\alpha}(u) \quad s.t. \ Ku = f \tag{6.19}$$

in case $\beta = 0$. Also, in both cases, $\lim_i R_{\alpha^{k_i}}(u^{k_i}) = \hat{R}_{\alpha}(\hat{u})$.

Proof. Given the properties we have obtained for R_{α} and the assumptions on S_{f^k}, S_f, the proof is now rather direct and we only provide a sketch for the sake of completeness.

At first note that, in case $\beta = 0$, existence of a solution $\hat{u}$ to (6.19) follows using Theorem 6.2 with $S_f = \mathcal{I}_{\{f\}}$, and assumption $ii)$ ensures a finite minimum. Further, since $\alpha_e^k \geq \alpha_e$ for all $e \in E$, which yields $\gamma_k = 1$ for γ_k according to (6.6), Theorem 6.6 implies that $R_{\alpha^k}(\hat{u}) \to \hat{R}_{\alpha}(\hat{u})$ and we get

$$S_{f^k}(Ku^k) + \beta_k R_{\alpha^k}(u^k) \leq \delta_k + \beta_k R_{\alpha^k}(\hat{u}) \to 0 \quad \text{as} \quad k \to \infty \tag{6.20}$$

using assumption $ii)$. Consequently, using hypothesis $i)$, it also holds that

$$R_{\alpha^k}(u^k) \leq \delta_k/\beta_k + R_{\alpha^k}(\hat{u}) \to \hat{R}_{\alpha}(\hat{u}) \quad \text{as} \quad n \to \infty. \tag{6.21}$$

This implies in particular boundedness of $S_{f^k}(Ku^k)$ and $R_{\alpha^k}(u^k)$.

In case $\beta > 0$, we can select $\hat{u}$ to be a solution to (6.18) and by Theorem 6.6 and convergence of S_{f^k} to S_f estimate according to

$$S_{f^k}(Ku^k) + \beta_k R_{\alpha^k}(u^k) \leq S_{f^k}(K\hat{u}) + \beta_k R_{\alpha^k}(\hat{u}) \to S_f(K\hat{u}) + \beta\hat{R}_{\alpha}(\hat{u}) \quad \text{as} \quad k \to \infty. \tag{6.22}$$

In particular, also in this case, both $S_{f^k}(Ku^k)$ and $R_{\alpha^k}(u^k)$ are bounded. Choosing Z as a complement of $\ker(K) \cap L$ in L, such that the projection $P_Z : L \to Z$ satisfies $\mathrm{rg}(\mathrm{Id} - P_Z) = \ker(K) \cap L$, and $\tilde{u}^k := u^k - P_L u^k + P_Z P_L u^k$, we observe that $\tilde{u}^k - u^k \in \ker(K) \cap L$ and, using equi-coercivity of $(S_{f^k})_k$ and Proposition 4.7, we can obtain, as in the proof of Lemma 4.2, that $(\tilde{u}^k)_k$ is bounded and hence admits a subsequence weakly converging in $X_{\hat{n}}$.

Now take $u \in X_{\hat{n}}$ to be the limit of a subsequence $(u^{k_i})_i$ of $(u^k)_k$ weakly converging in $X_{\hat{n}}$. In case $\beta = 0$, using weak lower semi-continuity, convergence of S_{f^k} to S_f, and that $S_f(v) = 0$ only if $v = f$, it follows from (6.20) and (6.21) that $Ku = f$ and $\hat{R}_{\alpha}(u) \leq \hat{R}_{\alpha}(\hat{u})$ as claimed in (6.19), and consequently, also that $\lim_i R_{\alpha^{k_i}}(u^{k_i}) = \hat{R}_{\alpha}(u)$. In case $\beta > 0$, again using weak lower semi-continuity and convergence of S_{f^k} to S_f, it follows from (6.22) and $\hat{R}_{\alpha}(u) \leq \liminf_i R_{\alpha^{k_i}}(u^{k_i})$ that u solves (6.18), and that $E_{k_i} := S_{f^{k_i}}(Ku^{k_i}) + \beta_{k_i} R_{\alpha^{k_i}}(u^{k_i}) \to E_0 := S_f(Ku) + \beta\hat{R}_{\alpha}(u)$. If $\limsup_i R_{\alpha^{k_i}}(u^{k_i}) > \hat{R}_{\alpha}(u)$, then the estimate

$$S_f(Ku) \leq \liminf_i \left(E_{k_i} - \beta_{k_i} R_{\alpha^{k_i}}(u^{k_i}) \right) = E_0 - \limsup_i R_{\alpha^{k_i}}(u^{k_i}) < S_f(Ku)$$

yields a contradiction, hence also $\lim_i R_{\alpha^{k_i}}(u^{k_i}) = \hat{R}_{\alpha}(u)$ and the proof is complete. $\square$

Remark 6.9. Theorem 6.8 is valid for several particular cases which are worth mentioning:

- If $\alpha > 0$ component-wise, then the above results hold for $\hat{R}_{\alpha} = R_{\alpha}$.

- If we fix $\alpha^k = \alpha$ and have $\beta = 0$, this is a classical convergence-for-vanishing-noise result for a fixed regularization functional.

- Regarding both β and α as regularization parameters, this is a rather general convergence result for multi-parameter regularization and we refer to [6, 41, 34] for related work.

- If we fix $f^k = f$, this is a stability result for varying the parameters α, β, which is in particular relevant in the context of bilevel optimization, see Section 7 below.

- Note that $\alpha_e^k \geq \alpha_e$ was only used in combination with Theorem 6.6 to ensure that $\lim_{k\to\infty} R_{\alpha^k}(u) = \hat{R}_\alpha(u)$. In case R_α is positively one-homogeneous, following Remark 6.7, this assumption can be dropped. Also, in case $f^k = f$ and $\beta > 0$, the assumption can be dropped in case S_f is continuous in the sense that $\lim_{\lambda\to 1} S_f(\lambda v) = S_f(v)$ for all $v \in \text{dom}(S_f)$.

- Again, as described in Remark 6.3, the result can be modified to hold without assuming reflexivity of $X_{\hat{n}}$.

7   Bilevel optimization

The goal of this section is to show well-posedness of a bilevel optimization problem for learning the weights α in a regularization graph. In order to allow for an arbitrary removal of different subtrees of the graph via setting $\alpha_e = 0$, we will need to include an additional penalty on the edge variables $(w_e)_{e \in E}$. To formulate this, we use the notation

$$R_\alpha(u, (w_e)_{e\in E}) := \sum_{n\in V} \Psi_n\Big(\Theta_{(n^-,n)} w_{(n^-,n)} - \sum_{(n,m)\in E} \alpha_{(n,m)} \Phi_{(n,m)} w_{(n,m)}\Big)$$

where again $w_{(\hat{n}^-,\hat{n})} = u$ and $\Theta_{(\hat{n}^-,\hat{n})} = \text{Id}$. Also, we need an assumption based on the invariant subspaces of regularization graph functionals. To formulate this, first recall the recursive representation of a regularization graph functional $R_\alpha = R(\mathcal{G}_\alpha)$ from Lemma 3.2 as

$$R_\alpha(u) = \inf\Big\{ \Psi_{\hat{n}}\big(u - \sum_{\hat{e}\in\hat{E}} \alpha_{\hat{e}}\Phi_{\hat{e}} w_{\hat{e}}\big) + \sum_{\hat{e}\in\hat{E}} R^{\hat{e}}_{\alpha^{\hat{e}}}(\Theta_{\hat{e}} w_{\hat{e}}) \;\Big|\; w_{\hat{e}} \in \text{dom}(\Theta_{\hat{e}}) \text{ for all } \hat{e}\in\hat{E}\Big\}. \quad (7.1)$$

Based on this, for $e \in E$, we henceforth denote $M^e := \Theta_e^{-1}(L^e)$, where Θ_e^{-1} is the inverse of $\Theta_e : \ker(\Theta_e)^\perp \to \text{rg}(\Theta_e)$ (recall that $\ker(\Theta_e)^\perp := \text{rg}(\text{Id} - P_{\ker(\Theta_e)}) \cap \text{dom}(\Theta_e)$ with $P_{\ker(\Theta_e)}$ according to Assumption (H6)) and L^e is the invariant subspace of the regularization graph functional $R^e_{\alpha^e} = R(\mathcal{G}^e_{\alpha^e})$ with $\mathcal{G}^e_{\alpha^e}$ the regularization graph corresponding to the subtree of G starting at edge $e \in E$ with functionals, spaces, operators and weights inherited from $\mathcal{G}_\alpha$. Note that M^e is finite dimensional by finite dimensionality of L^e and of $\ker(\Theta_e)$ for every $e \in E$. Finally, we denote the projection $P^e : \text{dom}(\Theta_e) \to M^e$ as

$$P^e w := \Theta_e^{-1} P_{\text{rg}(\Theta_e)\cap L^e} \Theta_e w + P^e_{\ker(\Theta_e)} w, \quad (7.2)$$

where $P^e_{\ker(\Theta_e)}$ is a projection onto $\ker(\Theta_e)$, noting that P^e is indeed a projection.
Using these notations, we now provide a lower semi-continuity result that includes vanishing weights as follows.

Lemma 7.1. *Let $\mathcal{G}_\alpha$ be a regularization graph with root node $\hat{n}$ and weights $\alpha \in [0,\infty)^E$, $R_\alpha = R(\mathcal{G}_\alpha)$ and $(\alpha^k)_k$ be a sequence of weights in $(0,\infty)^E$ such that $(\alpha^k)_k \to \alpha$.*

Then, with $(u^k)_k$ weak* converging to some $u \in X_{\hat{n}}$ and $((w_e^k)_{e \in E})_k$ a sequence realizing the minimum in (2.2) with u^k for u and α^k for α such that $(P^e w_e^k)_k$ and $(R_{\alpha^k}(u^k, ((w_e^k)_e)_k))_k$ are bounded, $((w_e^k)_{e \in E})_k$ is bounded and admits a subsequence converging weak* to some $(w_e)_{e \in E}$ such that

$$R_\alpha(u, (w_e)_e) \leq \liminf_{k \to \infty} R_{\alpha^k}(u^k, ((w_e^k)_e)_k). \tag{7.3}$$

Note that, in addition to explicitly including the variables $(w_e)_e$, this lower semi-continuity result differs from the one of Theorem 6.6 in the fact that in the limit, only the weights change (possibly to zero), but not the original regularization graph. This can be achieved thanks to the boundedness assumption on the sequences $(P^e w_e^k)_k$ that does not always hold true as clarified in the following remark.

Remark 7.2. Consider the regularization graph functional for TGV^2 (see Section 2.1) according to

$$R_{\alpha^k}(u^k) = \inf_{\substack{w_1 \in \mathrm{BV}(\Omega), \\ w_2 \in \mathcal{M}(\Omega, \mathbb{R}^d), \\ w_3 \in \mathrm{BD}(\Omega, \mathbb{R}^d)}} \mathcal{I}_{\{0\}}(u^k - w_1) + \mathcal{I}_{\{0\}}(\nabla w_1 - w_2 - \alpha_k w_3) + \|w_2\|_\mathcal{M} + \|\mathcal{E} w_3\|_\mathcal{M} \tag{7.4}$$

where $\alpha^k \to 0$ and $u^k = u$ for every k with $\nabla u \in \ker(\mathcal{E}) \setminus \{0\}$. Then, the sequence $(w_1^k, w_2^k, w_3^k)_k = (u, 0, \nabla u / \alpha^k)_k$ realizes the minimum in (7.4) with $R_{\alpha^k}(u^k) = 0$ for every k. However, for edge 3, we have $M^3 = \ker(\mathcal{E})$ and it holds that $\|P_{M^3}(\nabla u / \alpha^k)\|_{L^{d'}(\Omega, \mathbb{R}^d)} = (\alpha^k)^{-1} \|\nabla u\|_{L^{d'}(\Omega, \mathbb{R}^d)} \to +\infty$, showing that in this case, the assumptions of Lemma 7.1 do not hold.

Proof of Lemma 7.1. Again we proceed by induction, assuming the result holds true for all regularization graphs of height less than h and assume that the height of $\mathcal{G}_\alpha$ is h. The case $h = 0$ is again immediate and we assume $h \geq 1$. Writing

$$R_{\alpha^k}(u^k, (w_e^k)_e) = \Psi_{\hat{n}} \left(u^k - \sum_{\hat{e} \in \hat{E}} \alpha_{\hat{e}}^k \Phi_{\hat{e}} w_{\hat{e}}^k \right) + \sum_{\hat{e} \in \hat{E}} R_{(\alpha^k)^{\hat{e}}}^{\hat{e}} (\Theta_{\hat{e}} w_{\hat{e}}^k, (w_e^k)_{e \in E^{\hat{e}}})$$

with $\hat{E} \subset E$ the set all edges connected to the root node $\hat{n}$, $R_{(\alpha^k)^{\hat{e}}} = R(\mathcal{G}_{(\alpha^k)^{\hat{e}}}^{\hat{e}})$, and $(\mathcal{G}_{(\alpha^k)^{\hat{e}}}^{\hat{e}})$ regularization graphs of height less than h (see Lemma 3.2) and with root node $\hat{n}^{\hat{e}}$ we observe, estimating as in (6.12) and using boundedness of the $(\alpha_{\hat{e}}^k)_k$, for generic constants $C, D, \tilde{C} > 0$ independent of k, that

$$\|w_{\hat{e}}^k - P^{\hat{e}} w_{\hat{e}}^k\|_{X_{\hat{e}}} \leq C R_{(\alpha^k)^{\hat{e}}}^{\hat{e}} (\Theta_{\hat{e}} w_{\hat{e}}^k) + D < \tilde{C} < \infty,$$

hence boundedness of $(P^e w_e^k)_k$ for every $e \in E$ implies that $(w_{\hat{e}}^k)_k$ is bounded for $\hat{e} \in \hat{E}$. Further, again using the coercivity estimates for $R_{(\alpha^k)^{\hat{e}}}^{\hat{e}}$, the definition of $P^{\hat{e}}$, the estimate in (6.11) and the continuity of $\Theta_{\hat{e}} P^{\hat{e}}$ we obtain

$$\|\Theta_{\hat{e}} w_{\hat{e}}^k\|_{X_{\hat{n}^{\hat{e}}}} \leq \|\Theta_{\hat{e}} w_{\hat{e}}^k - \Theta_{\hat{e}} P^{\hat{e}} w_{\hat{e}}^k\|_{X_{\hat{n}^{\hat{e}}}} + \|\Theta_{\hat{e}} P^{\hat{e}} w_{\hat{e}}^k\|_{X_{\hat{n}^{\hat{e}}}}$$
$$\leq C R_{(\alpha^k)^{\hat{e}}}^{\hat{e}} (\Theta_{\hat{e}} w_{\hat{e}}^k) + D + \|\Theta_{\hat{e}} P^{\hat{e}} w_{\hat{e}}^k\|_{X_{\hat{n}^{\hat{e}}}} \leq \tilde{C} < +\infty,$$

where again $C, D, \tilde{C} > 0$ denote generic constants independent of k. Hence, by weak* compactness and weak* closedness of the $\Theta_{\hat{e}}$ we obtain that that $w_{\hat{e}}^k \overset{*}{\rightharpoonup} w_{\hat{e}} \in \mathrm{dom}(\Theta_{\hat{e}})$ as well as $\Theta_{\hat{e}} w_{\hat{e}}^k \overset{*}{\rightharpoonup} \Theta_{\hat{e}} w_{\hat{e}}$. The induction hypothesis together with the weak* lower semicontinuity of $\Psi_{\hat{n}}$ and the weak*-to-weak* continuity of $\Phi_{\hat{e}}$ implies then the result. $\square$

Consider now a regularization graph $\mathcal{G}_\alpha$ with root node $\hat{n}$ and let $R_\alpha = R(\mathcal{G}_\alpha) : X_{\hat{n}} \to [0, \infty]$ be the associated regularization functional. Let Z be a Banach space such that $Z \hookrightarrow X_{\hat{n}}$ and let H_1, H_2 be two functionals that penalize the weights α and auxiliary variables $(w_e)_{e \in E}$, respectively. We consider the bilevel optimization problem

$$
\inf_{\substack{\alpha \in [0,\infty)^E, \beta \in (0,\infty) \\ u_{\alpha,\beta} \in X_{\hat{n}}, (w_e^{\alpha,\beta})_{e \in E}}} \|u_{\alpha,\beta} - \hat{u}\|_Z + H_1(\alpha) + H_2((w_e^{\alpha,\beta})_{e \in E})
$$

$$
\text{s.t. } (u_{\alpha,\beta}, (w_e^{\alpha,\beta})_{e \in E}) \in \underset{u \in X_{\hat{n}}, (w_e)_{e \in E}}{\arg\min} \; S_f(Ku) + \beta R_\alpha(u, (w_e)_{e \in E}), \tag{7.5}
$$

where $\hat{u}$ is some ground truth datum and $f \approx K\hat{u}$ a corrupted measurement.

Remark 7.3. Note that this single-datum bilevel setting is a generic model problem for learning parameters from a larger training set $(\hat{u}_m, f_m)_m$. Indeed, the single-datum bilevel setting can be extended to a larger training set by simply vectorizing all involved quantities, for instance.

We now provide an existence result for the bilevel problem, where we use the convention that for $\beta = \infty$, we have $\beta R_\alpha(u, (w_e)_e) = 0$ if $R_\alpha(u, (w_e)_e) = 0$ and $\beta R_\alpha(u, (w_e)_e) = \infty$ else, and for which a concrete example and its assumptions are discussed after its proof below.

In this existence result, regarding the existence of an optimal parameter β, it is important to note that in (7.5), the parameter β is taken in the open interval $(0, \infty)$. This is necessary as otherwise, existence of a solution to the lower level problem cannot be guaranteed. The following theorem takes this into account by allowing the optimal parameter also to attain the value 0, in which case it states that existence to the lower level problem with $\beta = 0$ also holds, see Remark 7.5 for details.

Theorem 7.4. *Let Z be a Banach space, $\mathcal{G}_\alpha$ be a regularization graph with root node $\hat{n}$ and assume that $X_{\hat{n}}$ is reflexive with $Z \hookrightarrow X_{\hat{n}}$. Further, let $K : X_{\hat{n}} \to Y$ with Y a Banach space be linear and continuous, $S_f : Y \to [0, \infty]$ be a proper, convex, lower semi-continuous and coercive and $H_1 : [0, \infty)^E \to [0, \infty]$, $H_2 : \bigtimes_{e \in E} X_e \to [0, \infty]$ be proper and strongly and weak* lower semi-continuous functionals, respectively, such that $H_2(0) = 0$.*

Let $\hat{u} \in Z$ be a given ground truth variable such that S_f is the discrepancy with respect to a noisy version $f \in Y$ of $K\hat{u}$. Defining $\tilde{\alpha} \in [0, \infty)^E$ by $\tilde{\alpha}_e = \inf\{\alpha_e \mid \alpha \in \mathrm{dom}(H_1)\}$, further assume that

i) $\mathrm{dom}(R_{\tilde{\alpha}})$ is dense in $X_{\hat{n}}$,

ii) S_f is continuous and $\mathrm{dom}(S_f)$ open,

iii) $\mathrm{dom}(S_f \circ K) \cap L_{\tilde{\alpha}} \neq \emptyset$, where $L_{\tilde{\alpha}}$ is the invariant subspace of $R_{\tilde{\alpha}}$ as in Remark 4.4,

iv) $\|\alpha\| \leq CH_1(\alpha) + D$ for $C, D > 0$,

v) $\sum_{e \in E} \|P^e w_e\| \leq CH_2((w_e)_e) + D$ for $C, D > 0$ and with P^e as in (7.2).

Then, there exist $\hat{\alpha} \in [0, \infty)^E$, $\hat{\beta} \in [0, \infty]$, $u_{\hat{\alpha},\hat{\beta}}$ and $(w_e^{\hat{\alpha},\hat{\beta}})_{e \in E}$ such that $(u_{\hat{\alpha},\hat{\beta}}, (w_e^{\hat{\alpha},\hat{\beta}})_{e \in E})$ solves the lower level problem in (7.5) with parameters $(\hat{\alpha}, \hat{\beta})$ and such that

$$
\|u_{\hat{\alpha},\hat{\beta}} - \hat{u}\|_Z + H_1(\hat{\alpha}) + H_2((w_e^{\hat{\alpha},\hat{\beta}})_e) = \inf_{\substack{\alpha \in [0,\infty)^E, \beta \in (0,\infty), \\ u_{\alpha,\beta} \in X, (w_e^{\alpha,\beta})_e}} \|u_{\alpha,\beta} - \hat{u}\|_Z + H_1(\alpha) + H_2((w_e^{\alpha,\beta})_e)
$$

$$
\text{s.t. } (u_{\alpha,\beta}, (w_e^{\alpha,\beta})_e) \in \underset{u \in X_{\hat{n}}, (w_e)_e}{\arg\min} \; S_f(Ku) + \beta R_\alpha(u, (w_e)_e). \tag{7.6}
$$

Proof. In case the infimum in the bilevel problem (7.5) is infinite, any parameter combination together with a corresponding solution of the lower level problem will be a solution, hence we assume from now on that the infimum in (7.5) is finite. Take $(\alpha^k, \beta_k)_k$ to be a minimizing sequence in $[0, \infty)^E \times (0, \infty)$ for (7.5) with $(u^k)_k = (u_{\alpha^k, \beta_k})_k$ and $((w_e^k)_e)_k = ((w_e^{\alpha^k, \beta_k})_e)_k$ corresponding sequences of solutions to the lower level problem. Then, obviously $(u^k)_k$ is bounded in Z and by $\|\cdot\|_{X_{\hat{n}}} \le C \|\cdot\|_Z$ we obtain a (non-relabeled) subsequence weakly converging to some u in $X_{\hat{n}}$. By the coercivity of H_1 (hypothesis $iv)$) we can also assume that, up to a subsequence, $\alpha^k \to \hat{\alpha} \in \operatorname{dom}(H_1)$. By possibly considering another (non-relabeled) subsequence, we can further achieve that, for each $e \in E$, either $\alpha_e^k > 0$ for all k or $\alpha_e^k = 0$ for all k. Noting that in the latter case we can remove the subgraphs of $G = (V, E)$ after $e \in E$ with $\alpha_e^k = 0$ for all k without changing the value of R_α or any of the R_{α^k}, we can further assume that $\alpha_e^k > 0$ for all k and $e \in E$.

At first assume that there exists a subsequence of $(\beta_k)_k$ converging to zero. Then, moving to this subsequence, we obtain for any $z \in \operatorname{dom}(R_{\tilde{\alpha}}) \subset \operatorname{dom}(R_{\hat{\alpha}})$ (where $\operatorname{dom}(R_{\tilde{\alpha}}) \subset \operatorname{dom}(R_{\hat{\alpha}})$ follows from Lemma 3.3 and the definition of $\tilde{\alpha}$) and $z \in \operatorname{dom}(S_f \circ K)$ that

$$S_f(Ku) \le \liminf_k S_f(Ku^k) + \beta_k R_{\alpha^k}(u^k) \le \liminf_k S_f(\gamma_k Kz) + \beta_k R_{\alpha^k}(\gamma_k z)$$
$$\le \liminf_k S_f(\gamma_k Kz) + \beta_k R_{\hat{\alpha}}(z) = S_f(Kz),$$

where we have used that $R_{\alpha^k}(\gamma_k z) \le \hat{R}_{\hat{\alpha}}(z) \le R_{\hat{\alpha}}(z)$ for γ_k according to (6.6) by Theorem 6.6 and Lemma 6.5, and that S_f is continuous on $\operatorname{dom}(S_f)$ with $\operatorname{dom}(S_f)$ open (hypothesis $ii)$). Density of $\operatorname{dom}(R_{\tilde{\alpha}})$ and continuity of S_f then implies that $(u, (0)_{e \in E})$ is a solution to the lower level problem in (7.6) for $\hat{\beta} = 0$. Lower semi-continuity of $\|\cdot\|_Z$ and H_1, and the fact that $0 = H_2((0)_{e \in E}) \le \liminf_k H_2((w_e^k)_e)$ then yields the claimed optimality of $(\hat{\alpha}, \hat{\beta})$ with $\hat{\beta} = 0$.

Assume now that $(\beta_k)_k$ is unbounded such that, again by using a non-relabeled subsequence, we can assume that $\beta_k \to \infty$. Optimality and the estimate (6.8) then give for any $z \in \operatorname{dom}(S_f \circ K)$ with $R_{\hat{\alpha}}(z) = 0$ (such a z exists by hypothesis $iii)$ since $L_{\tilde{\alpha}} \subset L_{\hat{\alpha}}$, with $L_{\tilde{\alpha}}$ and $L_{\hat{\alpha}}$ being the invariant subspaces of $R_{\tilde{\alpha}}$ and $R_{\hat{\alpha}}$, respectively) that

$$S_f(Ku^k) + \beta_k R_{\alpha^k}(u^k, (w_e^k)_e) \le S_f(\gamma_k Kz) \to S_f(Kz) < \infty.$$

This implies in particular that $(R_{\alpha^k}(u^k, (w_e^k)_e))_k$ is bounded such that, using that $(P^e w_e^k)_k$ is bounded for each $e \in E$ due to coercivity of H_2 as in assumption $v)$, by Lemma 7.1, the sequence $((w_e^k)_e)_k$ admits a subsequence weak* converging to some $(w_e)_e$. Weak* lower semi-continuity then yields

$$R_{\hat{\alpha}}(u, (w_e)_e) \le \liminf_k R_{\alpha^k}(u^k, (w_e^k)_e) = 0.$$

Also, from weak lower semi-continuity of $S_f \circ K$, we obtain that $S_f(Ku) \le S_f(Kz)$ and, consequently, that $(u, (w_e)_e)$ solves the lower level problem in (7.6) for $(\hat{\alpha}, \hat{\beta})$ with $\hat{\beta} = \infty$. Lower semi-continuity of $\|\cdot\|_Z$ and H_1, and weak* lower semi-continuity of H_2 finally implies that $(\hat{\alpha}, \hat{\beta})$ is optimal as claimed.

At last assume that, again up to a non-relabeled subsequence, $\beta_k \to \hat{\beta} \in (0, \infty)$. Then, we get for any $z \in \operatorname{dom}(R_{\hat{\alpha}}) \cap \operatorname{dom}(S_f \circ K)$ (which again exists by hypothesis $iii)$ since $L_{\tilde{\alpha}} \subset \operatorname{dom}(R_{\tilde{\alpha}}) \subset \operatorname{dom}(R_{\hat{\alpha}})$) that

$$\liminf_k S_f(Ku^k) + \beta_k R_{\alpha^k}(u^k, (w_e^k)_e) \le \liminf_k S_f(\gamma_k Kz) + \beta_k R_{\alpha^k}(\gamma_k z) \le S_f(Kz) + \beta R_{\hat{\alpha}}(z),$$

such that again, $(R_{\alpha^k}(u^k, (w_e^k)_e))_k$ and $(P^e w_e^k)_k$ are bounded and by Lemma 7.1, we can assume that $((w_e^k)_e)_k$ admits a subsequence weak* converging to some $(w_e)_e$. Lower semi-continuity then yields

$$S_f(Ku) + \beta R_{\hat{\alpha}}(u, (w_e)_e) \leq S_f(Kz) + \beta R_{\hat{\alpha}}(z),$$

which shows that $(u, (w_e)_e)$ solves the lower level problem in (7.6). Finally, again lower semi-continuity of $\|\cdot\|_Z$, H_1 and weak* lower semi-continuity of H_2 imply optimality of $(\hat{\alpha}, \hat{\beta})$ as claimed. $\square$

Before discussing the assumptions and results of Theorem 7.4 in detail, we provide an example.

TGV^2-shearlet infimal convolution. With the notation of Section 2.1 we can define a regularization graph $\mathcal{G}_\alpha$ to be the one in Section 2.1 such that $R_\alpha = R(\mathcal{G}_\alpha) : L^2(\Omega) \to [0, \infty]$ with $\Omega \subset \mathbb{R}^2$ a bounded Lipschitz domain is given as

$$R_\alpha(u) = \inf_{\substack{w_1 \in \mathrm{BD}(\Omega), \\ w_2 \in L^2(\mathbb{R}^2)}} \|\nabla(u - \alpha_0 r_\Omega w_2) - \alpha_1 w_1\|_{\mathcal{M}} + \|\mathcal{E}w_1\|_{\mathcal{M}} + \|\mathcal{SH}w_2\|_1,$$

where $\alpha \in [0, \infty)^2$ is chosen accordingly. Then, for $K : L^2(\Omega) \to Y$ linear and continuous and some Banach space Y, we can consider the bilevel problem

$$\inf_{\substack{\beta \in (0, \infty), \\ (\alpha_0, \alpha_1) \in [0, c]^2}} \|u_{\alpha, \beta} - \hat{u}\|_{L_2(\Omega)} + \mathcal{I}_{[0, d]}\left(\left\|P_{\ker(\mathcal{E})} w_1^{\alpha, \beta}\right\|_{L^2(\Omega)}\right)$$

$$\text{s.t. } (u_{\alpha, \beta}, w_1^{\alpha, \beta}, w_2^{\alpha, \beta}) \in \underset{u, w_1, w_2}{\arg\min} \frac{1}{2}\|Ku - f\|_Y^2 + \beta R_\alpha(u, w_1, w_2), \tag{7.7}$$

where $c, d > 0$ and $P_{\ker(\mathcal{E})} : \mathrm{BD}(\Omega) \to \ker(\mathcal{E})$ is a projection to the finite dimensional space $\ker(\mathcal{E}) = \{x \mapsto Ax + b \,|\, A \in \mathbb{R}^{2 \times 2}, b \in \mathbb{R}^2, A^T = -A\}$. Then, with e_0 and e_1 being the edges such that $\alpha_{e_0} = \alpha_0$ and $\alpha_{e_1} = \alpha_1$, respectively, $H_1(\alpha) = \mathcal{I}_{[0,c]^2}((\alpha_{e_0}, \alpha_{e_1})) + \mathcal{I}_{\{\alpha \,|\, \alpha_e = 1 \text{ for } e \notin \{e_0, e_1\}\}}(\alpha)$ and $H_2(w_1, w_2) = \mathcal{I}_{[0,d]}(\|P_{\ker(\mathcal{E})} w_1\|_{L^2(\Omega)})$, we have with $\tilde{\alpha}_e = \inf\{\alpha_e \,|\, \alpha \in \mathrm{dom}(H_1)\}$ that $R_{\tilde{\alpha}} = \mathrm{TV}$. Hence, $\mathrm{dom}(R_{\tilde{\alpha}}) \supset C_c^\infty(\Omega)$ is dense in $X_{\hat{n}} = L^2(\Omega)$ and all assumptions of Theorem 7.4 hold.

Consequently there exist parameters $(\hat{\alpha}, \hat{\beta})$ which are optimal for (7.7) as stated in Theorem 7.4. Denoting by $\triangle$ the infimal convolution operator, we observe for the corresponding parameters $\hat{\alpha}_0, \hat{\alpha}_1$ that:

- If $\hat{\alpha}_0 = 0, \hat{\alpha}_1 > 0$, then $R_{\hat{\alpha}} = \mathrm{TGV}^2_{(1, 1/\hat{\alpha}_1)}$.

- If $\hat{\alpha}_0 = \hat{\alpha}_1 = 0$, then $R_{\hat{\alpha}} = \mathrm{TV}$.

- If $\hat{\alpha}_1 = 0, \hat{\alpha}_0 > 0$, then $R_{\hat{\alpha}} = \mathrm{TV} \triangle (\|\cdot\|_1 \circ \mathcal{SH})$.

- If $\hat{\alpha}_1 > 0, \hat{\alpha}_0 > 0$, then $R_{\hat{\alpha}} = \mathrm{TGV}^2_{(1, 1/\hat{\alpha}_1)} \triangle (\|\cdot\|_1 \circ \mathcal{SH})$.

Thus, the model is able to learn different functionals by modifying the graph accordingly. This extends directly, e.g., to learning the order of TGV or the infimal convolution of TGV with other regularization functionals. The term $\mathcal{I}_{[0,d]}\left(\|P_{\ker(\mathcal{E})} w_1^{\alpha, \beta}\|_{L^2(\Omega)}\right)$ puts a constraint on the norm of the projection of the auxiliary variable $w_1^{\alpha, \beta}$ to $\ker(\mathcal{E})$. Avoiding such a term

45

is also possible, but would lead to different limit functionals in case of vanishing α: Without a bound on the elements of $\ker(\mathcal{E})$, the limit graph in case $\hat{\alpha}_0 = \hat{\alpha}_1 = 0$ would in this example for instance be

$$\hat{R}_{\hat{\alpha}}(u) = \inf_{w \in \ker(\mathcal{E})} \|\nabla u - w\|_{\mathcal{M}}$$

instead of $R_{\hat{\alpha}}(u) = \|\nabla u\|_{\mathcal{M}}$. Hence, in case of using the infimal convolution of functionals with non-trivial invariant subspace, the limit functional still allows to subtract an arbitrary element of this subspace.

Remark 7.5. We now discuss necessity of the additional density and continuity assumptions of the theorem and the obtained result in more detail.

- If $\hat{\beta} = 0$, the theorem states that $u_{\hat{\alpha},\hat{\beta}}$ is a solution to

$$\min_{u \in X_{\hat{n}}} S_f(Ku),$$

  and in particular that a best approximation of the noisy data exists. Note that this is not true in general. In a classical Hilbert space setting with $S_f(v) = \|u - f\|_2^2$ for instance, existence of a best approximation for every $f \in Y$ is in fact equivalent to K having closed range [24].

- If $\hat{\beta} = \infty$, we see that $u_{\hat{\alpha},\hat{\beta}}$ solves

$$\min_{u \in X_{\hat{n}}} S_f(Ku) \quad \text{subject to } R_{\hat{\alpha}}(u) = 0.$$

  Here, it can be shown as in Theorem 6.2 that solution always exists and we could have alternatively used $\beta \in (0, \infty]$ in the bilevel problem (7.5).

- Density of $\mathrm{dom}(R_{\tilde{\alpha}})$ is only required in case $\hat{\beta} = 0$ to ensure optimality over the entire space instead of $\mathrm{dom}(R_{\hat{\alpha}})$. In particular, this assumption can be dropped by bounding the admissible β away from zero.

- The assumption S_f being continuous and $\mathrm{dom}(S_f)$ open is always fulfilled if, for instance, $S_f(u) = \|u - f\|_Y^q$. It can be replaced by the weaker assumption that $S_f(\gamma^k Kz) \to S_f(Kz)$ for all $z \in \mathrm{dom}(S_f \circ K)$ and $\gamma^k \in (0, 1]$ converging to 1 by either bounding β away from zero or reducing the optimality of $u_{\hat{\alpha},\hat{\beta}}$ in case $\hat{\beta} = 0$ to optimality with respect to all functions in $\mathrm{dom}(R_{\hat{\alpha}})$ instead of the entire space.

- The assumption $\mathrm{dom}(S_f \circ K) \cap L_{\tilde{\alpha}} \neq \emptyset$ is always true if, for instance, $0 \in \mathrm{dom}(S_f)$. It can be weakened to $\mathrm{dom}(S_f \circ K) \cap \mathrm{dom}(R_{\tilde{\alpha}}) \neq \emptyset$ if the set of admissible β is bounded above.

- Typical examples for H_1 fulfilling the assumption of Theorem 7.4 would be H_1 that constrains $\alpha_e \in [0, c]$ for all $e \in E_0$ or penalizes $\sum_{e \in E_0} |\alpha_e|$, with some $E_0 \subset E$, and fixes $\alpha_e = 1$ for all remaining $e \in E \setminus E_0$. Here, a penalization of $\sum_{e \in E_0} |\alpha_e|$ is expected to promote sparsity of α and hence, a reduced complexity of the optimal regularization graph. The purpose of the constraints $\alpha_e = 1$ for $e \in E \setminus E_0$ is to avoid overparametrization, i.e., the usage of unnecessary parameters. This happens, for instance, in case of splitting nodes, i.e., if $\Psi_n = \mathcal{I}_{\{0\}}$ for some n. Further, the constraint $\alpha_e = 1$ for $e \in E \setminus E_0$ can be used to avoid $R_{\tilde{\alpha}} = \mathcal{I}_{\{0\}}$, which is the case if all weights are set to zero and $\Psi_{\hat{n}} = \mathcal{I}_{\{0\}}$ (see Definition 2.3).

- The coercivity of H_2 is only required on the finite dimensional spaces M_e for all $e \in E$, and is used to allow for the bilevel framework to cut edges of the graph by setting weights to zero. Without this assumption, a similar existence result with R_α being replaced by $\hat{R}_\alpha$ can be obtained.

8 Conclusions

In this work, we have introduced *regularization graphs* as a flexible framework for designing regularization functionals for the variational regularization of inverse problems. The proposed framework thoroughly covers existing regularization approaches and allows to define new ones in a simple and constructive way, essentially by drawing corresponding regularization graphs. We have provided a comprehensive analysis of the class of functionals derived from regularization graphs, which in particular includes well-posedness and convergence results for applying this class of functionals in a general inverse problems setting. Furthermore, we have developed and analyzed a bilevel optimization approach that allows to learn an optimal structure and complexity of a regularization graph, and hence of the corresponding regularization functional, from training data.
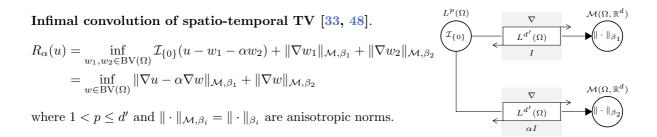
Future goals are to develop an equally flexible numerical framework for the application of regulariazation graphs to general inverse problems, as well as the numerical realization of the proposed bilevel approach.

A Appendix

Here we provide a list extending the examples of Section 2.1, that outlines the representation of different, existing regularization functionals as regularization graphs. Note that, as discussed in Section 3, also a finite combination of any of those functionals via summation or infimal convolution can again be represented as regularization graph.
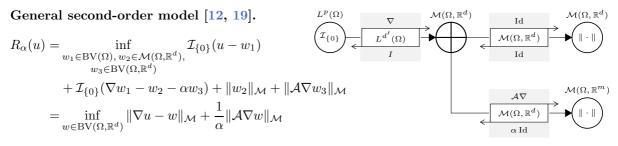
TV $-L^q$ infimal convolution [14, 15].

$$R_\alpha(u) = \inf_{\substack{w_1 \in BV(\Omega),\, w_2 \in \mathcal{M}(\Omega,\mathbb{R}^d), \\ w_3 \in L^q(\Omega,\mathbb{R}^d)}} \mathcal{I}_{\{0\}}(u - w_1)$$

$$+ \mathcal{I}_{\{0\}}(\nabla w_1 - w_2 - \alpha w_3) + \|w_2\|_{\mathcal{M}} + \|w_3\|_{L^q}$$

$$= \inf_{w \in L^q(\Omega,\mathbb{R}^d)} \|\nabla u - \alpha w\|_{\mathcal{M}} + \|w\|_{L^q}$$

with $1 < p \leq d'$ and $q \in (1, \infty)$.

Infimal convolution of spatio-temporal TV [33, 48].

$$R_\alpha(u) = \inf_{w_1, w_2 \in BV(\Omega)} \mathcal{I}_{\{0\}}(u - w_1 - \alpha w_2) + \|\nabla w_1\|_{\mathcal{M},\beta_1} + \|\nabla w_2\|_{\mathcal{M},\beta_2}$$

$$= \inf_{w \in BV(\Omega)} \|\nabla u - \alpha \nabla w\|_{\mathcal{M},\beta_1} + \|\nabla w\|_{\mathcal{M},\beta_2}$$

where $1 < p \leq d'$ and $\|\cdot\|_{\mathcal{M},\beta_i} = \|\cdot\|_{\beta_i}$ are anisotropic norms.

Sum of convex functions of TV **and** TV^2 **[42].**

$$R_\alpha(u) = \inf_{w \in \mathrm{BV}^2(\Omega)} \mathcal{I}_{\{0\}}(u-w) + \Psi_{f,g}\big((\nabla w, \nabla^2 w)\big)$$
$$= \|\nabla u\|_f + \|\nabla^2 u\|_g$$

where $1 < p \le d'$, $\Psi_{f,g}(z_1, z_2) = \|f(z_1)\|_{\mathcal{M}} + \|g(z_2)\|_{\mathcal{M}}$ with $\|f(\cdot)\|_{\mat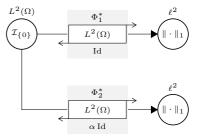hcal{M}}$, $\|g(\cdot)\|_{\mathcal{M}}$ appropriate convex functions of measures generalizing $\|\cdot\|_{\mathcal{M}}$ such that coercivity and weak* lower semicontinuity holds.
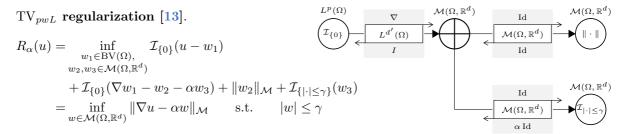
General second-order model [12, 19].

$$R_\alpha(u) = \inf_{\substack{w_1 \in \mathrm{BV}(\Omega),\, w_2 \in \mathcal{M}(\Omega, \mathbb{R}^d), \\ w_3 \in \mathrm{BV}(\Omega, \mathbb{R}^d)}} \mathcal{I}_{\{0\}}(u - w_1)$$
$$+ \mathcal{I}_{\{0\}}(\nabla w_1 - w_2 - \alpha w_3) + \|w_2\|_{\mathcal{M}} + \|\mathcal{A}\nabla w_3\|_{\mathcal{M}}$$
$$= \inf_{w \in \mathrm{BV}(\Omega, \mathbb{R}^d)} \|\nabla u - w\|_{\mathcal{M}} + \frac{1}{\alpha}\|\mathcal{A}\nabla w\|_{\mathcal{M}}$$

where $1 < p \le d'$, $m \in \mathbb{N}$ and the linear operator $\mathcal{A} : \mathbb{R}^{d \times d} \to \mathbb{R}^m$ is defined pointwise on ∇w such that suitable lower semicontinuity and coercivity assumptions hold.

Infimal convolution of tight frames [35].

$$R_\alpha(u) = \inf_{w_1, w_2 \in L^2(\Omega)} \mathcal{I}_{\{0\}}(u - w_1 - \alpha w_2) + \|\Phi_1^* w_1\|_1 + \|\Phi_2^* w_2\|_1$$
$$= \inf_{u = w_1 + w_2} \|\Phi_1^* w_1\|_1 + \frac{1}{\alpha}\|\Phi_2^* w_2\|_1$$

where Φ_i^* are associated with tight frames such as curvelets or Gabor frames [35] and $\|\cdot\|_1$ is the extension to $+\infty$ of the ℓ^1-norm to ℓ^2.

TV_{pwL} regularization [13].

$$R_\alpha(u) = \inf_{\substack{w_1 \in \mathrm{BV}(\Omega), \\ w_2, w_3 \in \mathcal{M}(\Omega, \mathbb{R}^d)}} \mathcal{I}_{\{0\}}(u - w_1)$$
$$+ \mathcal{I}_{\{0\}}(\nabla w_1 - w_2 - \alpha w_3) + \|w_2\|_{\mathcal{M}} + \mathcal{I}_{\{|\cdot| \le \gamma\}}(w_3)$$
$$= \inf_{w \in \mathcal{M}(\Omega, \mathbb{R}^d)} \|\nabla u - \alpha w\|_{\mathcal{M}} \quad \text{s.t.} \quad |w| \le \gamma$$

where $|w| \in \mathcal{M}^+(\Omega)$ is the variation of the measure $w \in \mathcal{M}(\Omega, \mathbb{R}^d)$, $\gamma \in \mathcal{M}^+(\Omega)$ is a given positive measure and $|w| \le \gamma$ means that $\gamma - |w|$ is a positive measure.

For the sake of completeness, we also provide the proof of the equivalence of a coercivity- and and closed-range assertion for the operators considered in this paper.

Lemma A.1. *Let $\Theta : \mathrm{dom}(\Theta) \subset X_e \to X_m$ be a linear operator between Banach spaces X_e and X_m that both admit a predual space and such that bounded sequences in X_e admit weak* convergent subsequences. Further, assume that Θ is weak* closed and has finite dimensional*

kernel. Then, there exists $C > 0$ and $P_{\ker(\Theta)} : X_e \to \ker(\Theta)$ a linear, continuous projection such that

$$\|w - P_{\ker(\Theta)}w\|_{X_e} \leq C\|\Theta w\|_{X_m}$$

for all $w \in \mathrm{dom}(\Theta)$ if and only if Θ has closed range.

Proof. Assuming that the coercivity assertion holds, the closedness of $\mathrm{rg}(\Theta)$ can be proven directly using the weak* closedness of Θ. On other hand, if $\mathrm{rg}(\Theta)$ is closed, then from [11, Remark 2.18] we deduce that there exists $\tilde{C} > 0$ such that

$$\inf_{z \in \ker(\Theta)} \|w - z\|_{X_e} \leq \tilde{C}\|\Theta w\|_{X_m} \quad \forall w \in \mathrm{dom}(\Theta). \tag{A.1}$$

In particular, due to finite dimensionality of $\ker(\Theta)$, we deduce the existence of a map $G : \mathrm{dom}(\Theta) \to \ker(\Theta)$ such that

$$\|w - G(w)\|_{X_e} \leq \tilde{C}\|\Theta w\|_{X_m} \quad \forall w \in \mathrm{dom}(\Theta). \tag{A.2}$$

Defining now

$$R(w) := \begin{cases} \|\Theta w\|_{X_m} & w \in \mathrm{dom}(\Theta), \\ +\infty & \text{otherwise,} \end{cases} \tag{A.3}$$

and arbitrarily extending G outside $\mathrm{dom}(\Theta)$ to a function $G : X_e \to \ker(\Theta)$ we obtain that

$$\|w - G(w)\|_{X_e} \leq \tilde{C}R(w) \quad \forall w \in X_e. \tag{A.4}$$

Finally, applying Lemma 4.1 with $D = 0$ (which yields $\tilde{D} = 0$) and $K = X_e$, we obtain the existence of a bounded, linear projection $P_{\ker(\Theta)}$ and a constant $C > 0$ such that the claimed coercivity holds. $\qquad\square$

References

[1] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems.* Oxford University Press, 2000.

[2] Hedy Attouch and Haim Brezis. *Duality for the sum of convex functions in general Banach spaces,* volume 34 of *Aspects of Mathematics and its Applications.* Elsevier, 1986.

[3] Gilles Aubert and Pierre Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations.* Springer, 2006.

[4] Jonathan M. Borwein and Jon D. Vanderwerff. *Convex functions: constructions, characterizations and counterexamples,* volume 109 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, 2010.

[5] Kristian Bredies. Symmetric tensor fields of bounded deformation. *Ann. di Mat. Pura ed Appl.,* 192(5):815–851, 2013.

[6] Kristian Bredies and Martin Holler. Regularization of linear inverse problems with total generalized variation. *J. Inv. Ill-Posed Problems,* 22(6):871–913, 2014.

[7] Kristian Bredies and Martin Holler. Higher-order total variation approaches and generalisations. *Inverse Probl.,* 36(12):123001, 2020.

[8] Kristian Bredies, Karl Kunisch, and Thomas Pock. Total generalized variation. *SIAM J. Imaging Sci.,* 3(3):492–526, 2010.

[9] Kristian Bredies and Dirk Lorenz. *Mathematical Image Processing*. Springer, 2018.

[10] Kristian Bredies and Hanna Katriina Pikkarainen. Inverse problems in spaces of measures. *ESAIM Control Optim. Calc. Var.*, 19(1):190–218, 2013.

[11] Haim Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011.

[12] Eva-Maria Brinkmann, Martin Burger, and Joana Sarah Grah. Unified models for second-order TV-type regularisation in imaging: A new perspective based on vector operators. *J. Math. Imaging Vis.*, 61(5):571–601, 2019.

[13] Martin Burger, Yury Korolev, Carola-Bibiane Schönlieb, and Christiane Stollenwerk. A total variation based regularizer promoting piecewise-Lipschitz reconstructions. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 485–497. Springer, 2019.

[14] Martin Burger, Konstantinos Papafitsoros, Evangelos Papoutsellis, and Carola-Bibiane Schönlieb. Infimal convolution regularisation functionals of BV and L^p spaces. The case $p = \infty$. In *IFIP Conference on System Modeling and Optimization*, pages 169–179. Springer, 2015.

[15] Martin Burger, Konstantinos Papafitsoros, Evangelos Papoutsellis, and Carola-Bibiane Schönlieb. Infimal convolution regularisation functionals of BV and L^p spaces. Part I: The finite p case. *J. Math. Imaging Vis.*, 55:343, 2016.

[16] Luca Calatroni, Chung Cao, Juan Carlos De los Reyes, Carola-Bibiane Schönlieb, and Tuomo Valkonen. Bilevel approaches for learning of variational imaging models. In *Variational methods*, volume 18 of *Radon Ser. Comput. Appl. Math.*, pages 252–290. De Gruyter, Berlin, 2017.

[17] Antonin Chambolle, Martin Holler, and Thomas Pock. A convex variational model for learning convolutional image atoms from incomplete data. *J. Math. Imaging Vis.: Special Issue on Mathematics of Deep Learning*, 62:417–444, 2020.

[18] Antonin Chambolle and Pierre-Louis Lions. Image recovery via total variation minimization and related problems. *Numer. Math.*, 76(2):167–188, 1997.

[19] Tony F. Chan, Selim Esedoglu, and Frederick E. Park. A fourth order dual method for staircase reduction in texture extraction and image restoration problems. *IEEE Int. Conf. Image Process.*, pages 4137–4140, 2010.

[20] Elisa Davoli, Irene Fonseca, and Pan Liu. Adaptive image processing: first order PDE constraint regularizers and a bilevel training scheme. *arXiv preprint arXiv:1902.01122*, 2019.

[21] J. C. De los Reyes, C.-B. Schönlieb, and T. Valkonen. Bilevel parameter learning for higher-order total variation regularisation models. *J. Math. Imaging Vision*, 57(1):1–25, 2017.

[22] Ivar Ekeland and Roger Témam. *Convex analysis and variational problems*, volume 28 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999.

[23] Michael Elad. *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer Science & Business Media, 2010.

[24] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and Its Applications*. Springer Science & Business Media, 1996.

[25] V. Estellers, S. Soatto, and X. Bresson. Adaptive regularization with the structure tensor. *IEEE Trans. Image Process.*, 24(6), 2015.

[26] F. Frühauf, C. Pontow, and O. Scherzer. Texture enhancing based on variational image decomposition. In *Mathematical Image Processing*, volume 5 of *Springer Proceedings in Mathematics*, pages 127–140. Springer, 2011.

[27] Yiming Gao and Kristian Bredies. Infimal convolution of oscillation total generalized variation for the recovery of images with structured texture. *SIAM J. Imaging Sci.*, 11(3):2021–2063, 2018.

[28] Yiming Gao and Xiaoping Yang. Infimal convolution type regularization of TGV and shearlet transform for image restoration. *Comput. Vis. Image Underst.*, 182:38–49, 2019.

[29] Markus Grasmair, Markus Haltmeier, and Otmar Scherzer. The residual method for regularizing ill-posed problems. *Appl. Math. Comput.*, 218(6):2693–2710, 2011.

[30] W. Guo, J. Qin, and W. Yin. A new detail-preserving regularization scheme. *SIAM J. Imaging Sci.*, 7(2):1309–1334, 2014.

[31] W. Hinterberger and O. Scherzer. Variational methods on the space of functions of bounded Hessian for convexification and denoising. *Computing*, 76(1-2):109–133, 2006.

[32] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer. A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Probl.*, 23(3):987–1010, 2007.

[33] Martin Holler and Karl Kunisch. On infimal convolution of TV-type functionals and applications to video and image reconstruction. *SIAM J. Imaging Sci.*, 7(4):2258–2300, 2014.

[34] Kazufumi Ito, Bangti Jin, and Tomoya Takeuchi. Multi-parameter Tikhonov regularization. *Methods and Applications of Analysis*, 18(1):31–46, 2011.

[35] G. Kutyniok. Clustered sparsity and separation of cartoon and texture. *SIAM J. Imaging Sci.*, 6(2):848–874, 2013.

[36] Gitta Kutyniok and Demetrio Labate, editors. *Shearlets*. Birkhäuser, 2012.

[37] Marc Lebrun, Miguel Colom, Antoni Buades, and Jean-Michel Morel. Secrets of image denoising cuisine. *Acta Numer.*, 21:475–576, 2012.

[38] Stamatios Lefkimmiatis, Anastasios Roussos, Petros Maragos, and Michael Unser. Structure tensor total variation. *SIAM J. Imaging Sci.*, 8:1090–1122, 2015.

[39] S. Mallat. *A wavelet tour of signal processing*. Elsevier, 2009.

[40] Yves Meyer. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*. American Mathematical Society, USA, 2001.

[41] Valeriya Naumova and Sergei V Pereverzyev. Multi-penalty regularization with a component-wise penalization. *Inverse Probl.*, 29(7):075002, 2013.

[42] K. Papafitsoros and C.-B. Schönlieb. A combined first and second order variational approach for image reconstruction. *J. Math. Imaging Vis.*, 48(2):308–338, 2014.

[43] Simone Parisotto, Jan Lellmann, Simon Masnou, and Carola-Bibiane Schönlieb. Higher-order total directional variation: Imaging applications. *SIAM J. Imaging Sci.*, 13(4):2063–2104, 2020.

[44] Simone Parisotto, Simon Masnou, and Carola-Bibiane Schönlieb. Higher-order total directional variation: Analysis. *SIAM J. Imaging Sci.*, 13(1):474–496, 2020.

[45] R. Tyrrell Rockafellar. *Conjugate duality and optimization*. Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1974.

[46] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, 1992.

[47] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*. Springer, 2009.

[48] Matthias Schloegl, Martin Holler, Andreas Schwarzl, Kristian Bredies, and Rudolf Stollberger. Infimal convolution of total generalized variation functionals for dynamic MRI. *Magn. Reson. Med.*, 78(1):142–155, 2017.

[49] Jean-Luc Starck, Emmanuel J Candès, and David L Donoho. The curvelet transform for image denoising. *IEEE Trans. Image Process.*, 11(6):670–684, 2002.

[50] A. N. Tihonov. On the solution of ill-posed problems and the method of regularization. *Dokl. Akad. Nauk SSSR*, 151:501–504, 1963.