Westfälische Wilhelms-Universität Münster

Nonlinear Spectral Decomposition

Martin Burger

Cells in Motion
Cluster of Excellence

Applied Mathematics Münster

Joint work with

Martin Benning (Münster / Cambridge)
MB², Meth. Appl. Anal. 2013, Special Issue Osher 70
Benning, PhD thesis, 2011

Guy Gilboa (Technion), Michael Möller (Munich), Lina Eckardt (Münster), Daniel Cremers (Munich)
SSVM 2015 / BsC Thesis 2014
Preprint (arXiV) 2016

Singular Value Decomposition

Singular values and singular vectors are crucial for the analysis of linear methods for solving inverse problems

$$Ku = f$$

Singular vectors are obtained as solutions of eigenvalue problem

$$\lambda K^* K u_\lambda = u_\lambda$$

Singular value $\sigma = \dfrac{1}{\sqrt{\lambda}}$

Convex Variational Regularization

In the last years linear reconstruction methods lost importance

Popular approaches (in particular in imaging) are of the form

$$\hat{u} \in \arg\min_{u \in \mathrm{dom}(J)} \left\{ \frac{1}{2} \|Ku - f\|_{\mathcal{H}}^2 + \alpha J(u) \right\}$$

with one-homogeneous J like TV or L1

Are there singular vectors for such ? Are they useful ?

Defining Singular Vectors

We need variational characterization for comparison

Note: linear case corresponds to

$$J(u) = \frac{1}{2} \|u\|^2$$

Rayleigh-principle: singular vector for smallest singular value minimizes

$$\|u\| \text{ subject to } \|Ku\| = 1$$

Ground States

Generalize Rayleigh principle:

$$u_0 \in \arg\min_{\substack{u \in \mathrm{dom}(J) \\ \|Ku\|_{\mathcal{H}}=1}} \{J(u)\}$$

Problem: can yield uninteresting elements minimizing *J*

Example *J=TV*: Ground states would simply be constant functions

Ground States

Choose J to be a seminorm on a dense subspace.
Then its kernel is a closed linear subspace

Eliminate kernel for improved definition of ground state

$$u_0 \in \arg\min_{\substack{u \in \ker(J)^\perp \\ \|Ku\|_{\mathcal{H}}=1}} \{J(u)\}$$

$$\lambda_0 = J(u_0)$$

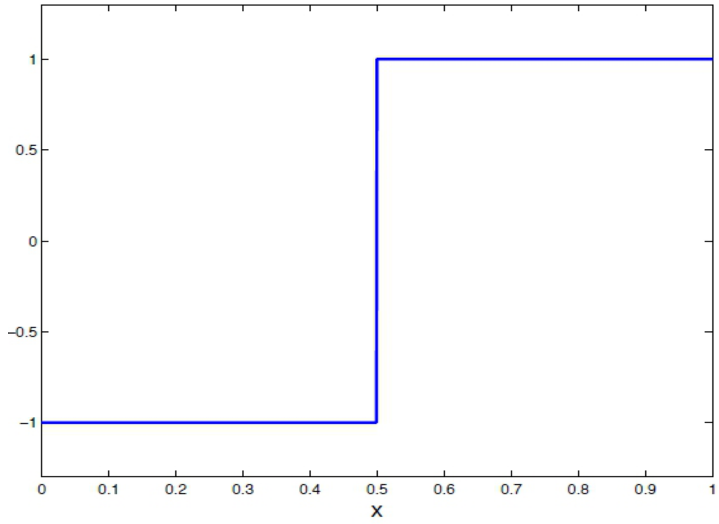
Existence under standard assumptions

Martin Burger

Ground States: Examples

1D Total Variation denoising ($K=I$):

ground state = single step function



Ground States: Examples

Sparsity ($J(u) = \|u\|_{\ell^1}$):
ground state = vector with nonzero entry at index corresponding to column of *K* with maximal norm

Nuclear norm of matrices:
ground state = rank one matrix corresponding to classical largest singular value

Ground States and Singular Vectors

Ground states are stationary points of Lagrangian

$$L(u;\lambda) = J(u) - \frac{\lambda}{2}(\|Ku\|_{\mathcal{H}}^2 - 1)$$

Due to nonconvexity of constraint there are multiple stationary points satisfying

$$\lambda K^* K u_\lambda \in \partial J\left(u_\lambda\right)$$

We call them singular vectors and focus on those with

$$\|Ku_\lambda\|_{\mathcal{H}} = 1$$

Rayleigh Principle for Higher Singular Vectors

Usual construction for further singular vectors

$$u_{\lambda_{n+1}} \in \arg\min_{u \in \mathcal{C}_n} J(u)$$

Due to nonconvexity of constraint there are multiple stationary points satisfying

$$\mathcal{C}_n := \{\ u \in \mathcal{U}\ \ \big|\ \|u\|_K = 1, \langle u_{\lambda_j}, u \rangle_K = 0, j = 0, \ldots, n\ \}$$

where

$$\langle u, v \rangle_K := \langle Ku, Kv \rangle$$

Rayleigh Principle for Singular Vectors

Usual construction for further singular vectors fails !

Using Lagrange multipliers we find

$$\lambda K^* K u_{\lambda_{n+1}} + \sum_{j=0}^{n} \mu_j K^* K u_{\lambda_j} = p_{\lambda_{n+1}} \in \partial J(u_{\lambda_{n+1}})$$

with

$$\mu_k = \langle p_{\lambda_{n+1}}, u_{\lambda_k} \rangle_{\mathcal{U}}$$

No particular reason for those to vanish !

Rayleigh Principle for Higher Singular Vectors

Construction for special cases can still be interesting

1D total variation denoising with appropriate TV definition:

$$\mathrm{TV}_*(u) := \sup_{\substack{\varphi \in C^\infty(\Omega;\mathbb{R}^n) \\ \|\varphi\|_{L^\infty(\Omega;\mathbb{R}^n)} \leq 1}} \int_\Omega u \,\mathrm{div}\varphi \, dx$$

$$\mathrm{TV}_*(u) = \int_0^1 |u'(x)| \; dx + |u(1)| + |u(0)|$$

Rayleigh principle yields sequence of singular vectors equivalent to the Haar wavelet basis !

Use of Singular Vectors ?

Due to nonlinearity, there is no singular value decomposition

Other ways of use:

- Canonical cases and exact solutions for regularization methods, analysis of bias
- Definition of scale relative to regularization, scale estimates

Exact Solutions of Variational Regularization

Solutions of

$$\hat{u} \in \arg\min_{u \in \mathrm{dom}(J)} \left\{ \frac{1}{2} \|Ku - f\|_{\mathcal{H}}^2 + \alpha J(u) \right\}$$

with $f = \gamma K u_\lambda$

are given by $\hat{u} = c u_\lambda$

$$c = \gamma - \alpha\lambda \qquad \gamma > \alpha\lambda$$

Similar results for noisy perturbations

Exact Solutions of Inverse Scale Space Method

Solutions of

$$\partial_t p(t) = K^*(f - Ku(t)), \qquad p(t) \in \partial J(u(t))$$

with $f = \gamma K u_\lambda$

are given by

$$u(t) = \begin{cases} 0 & if\, t < t_* = \lambda/\gamma \\ \gamma u_\lambda & if\, t \geq t_* = \lambda/\gamma \end{cases}$$

Similar results for noisy data and other related methods

Exact Solutions of Variational Regularization

Provides systematic way of analyzing exact solutions

Includes all examples in literature (most being ground states, some already charaterized as eigenfunctions):

- TV: *Strong-Chan 1996, Meyer 2001, Strong 2003*
- TV-flow: *Bellettini et al 2001*, *Andreu et al 2001*, *Caselles-Chambolle-Novaga 2007-2010*

- Higher order TV: *Papafitsoros-Bredies 2014, Pöschl-Scherzer 2014* mainly contained as singular values in *Benning-Brune-mb-Müller 2013*

Higher order TV functionals avoiding staircasing

Major idea: combine TV with higher order TV

Infimal convolution

$$\mathrm{ICTV}_\beta(u) := (\mathrm{TV} \square \beta \mathrm{TV}^2)(u) = \inf_{u=v+w} \left(\mathrm{TV}(v) + \beta\, \mathrm{TV}^2(w) \right)$$

Dual version

$$\mathrm{ICTV}_\beta(u) = \sup_{\substack{p \in C_0^\infty(\Omega;\mathbb{R}^d) \\ q \in C_0^\infty(\Omega;\mathrm{Sym}^2(\mathbb{R}^d)) \\ \|p\|_\infty \leq 1,\ \|q\|_\infty \leq 1 \\ \beta\, \mathrm{div}^2(q) = \mathrm{div}(p)}} \int_\Omega u\, \mathrm{div}(p)\, dx$$

Chambolle-Lions 97

TGV / GTV

Decomposition by inf-convolution not optimal, improvement by stronger dual constraint

$$\mathrm{GTV}_\beta(u) := \sup_{\substack{q \in C_0^\infty(\Omega;\mathrm{Sym}^2(\mathbb{R}^d)) \\ \|q\|_\infty \leq \beta,\ \|\mathrm{div}(q)\|_\infty \leq 1}} \int_\Omega u \ \mathrm{div}^2(q) \ dx$$

Primal version

$$\mathrm{GTV}_\beta(u) = \inf_{\substack{u=v+w \\ s \in \mathcal{N}(\mathrm{div})}} \left\{ \int_\Omega |\nabla v - s| + \beta \int_\Omega |\nabla^2 w + \nabla s| \right\}$$

Bredies et al 2011

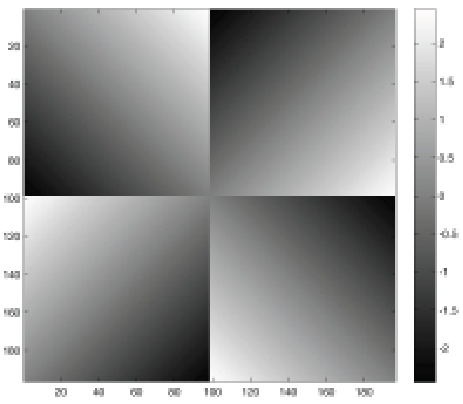
TGV vs. ICTV

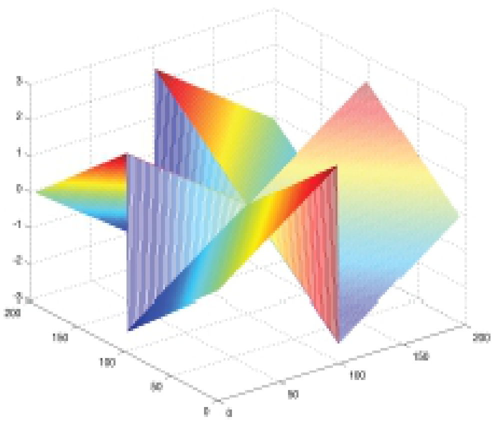
Equivalence of functionals in 1D

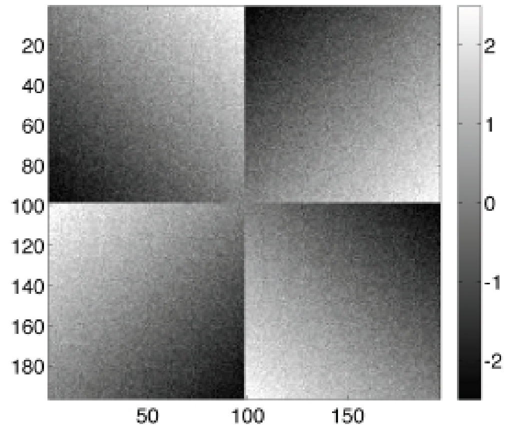
Intuitive advantages of TGV in multiple dimensions
Bredies et al 2011 / 2013, Benning-Brune-mb-Müller 2013

Better understanding by constructing eigenfunctions for TGV denoising, which are not eigenfunctions of ICTV .
Any eigenfunction of ICTV is eigenfunction of TGV
Müller PhD 2013

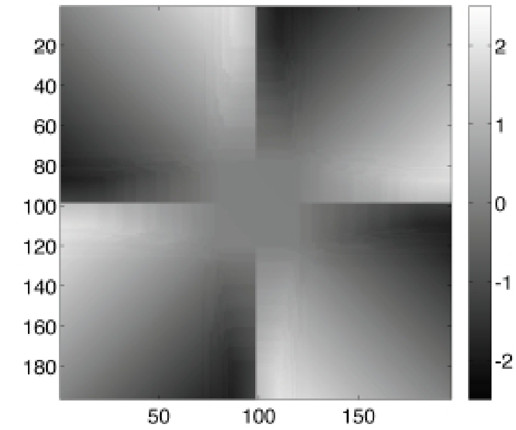
GTV Origami



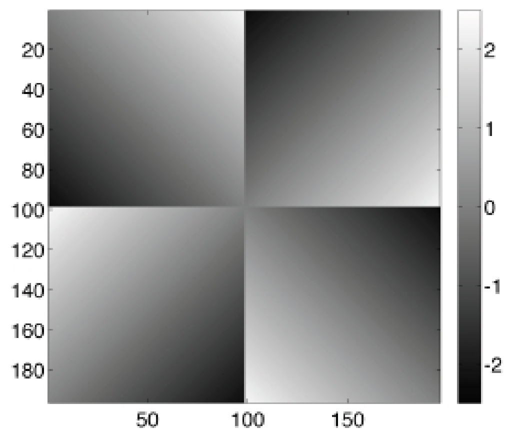




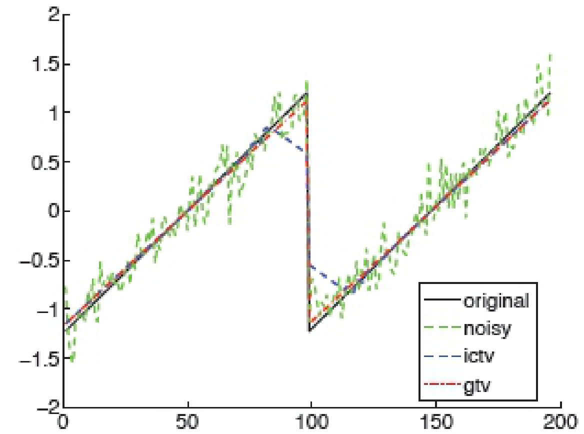
(a) noisy



(b) ICTV



(c) GTV



(d) line profile, line 50

Bias of Variational Regularization

Error (bias) in solution increasing with size of singular value

Does the smallest singular value define minimal bias ?

Arbitrary data satisfying only $\Vert f \Vert_{\mathcal{H}} \geq \alpha \lambda_0$

Then

$$\Vert K u^\alpha - f \Vert_{\mathcal{H}} \geq \alpha \lambda_0$$

where

$$u^\alpha \in \arg\min_{u \in \mathcal{U}} \left(\frac{1}{2} \Vert K u - f \Vert_{\mathcal{H}}^2 + \alpha J(u) \right)$$

Bias of Variational Regularization

In the same way underestimation of regularization functional

$$f = K\tilde{u} \ \ for \ \tilde{u} \in \mathcal{U} \ \ with \ J(\tilde{u}) < \infty$$

then

$$J(u^\alpha) \leq J(\tilde{u}) - \frac{\alpha}{2}\lambda_0^2$$

Spectral Decomposition

Consider simpler case of K = Id (eigenfunctions / values)

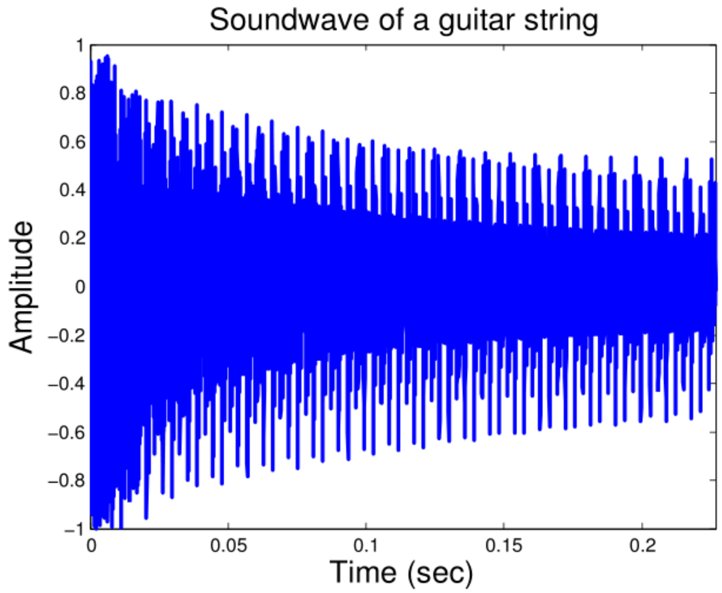
Can we get a spectral decomposition from a seminorm *J* ?

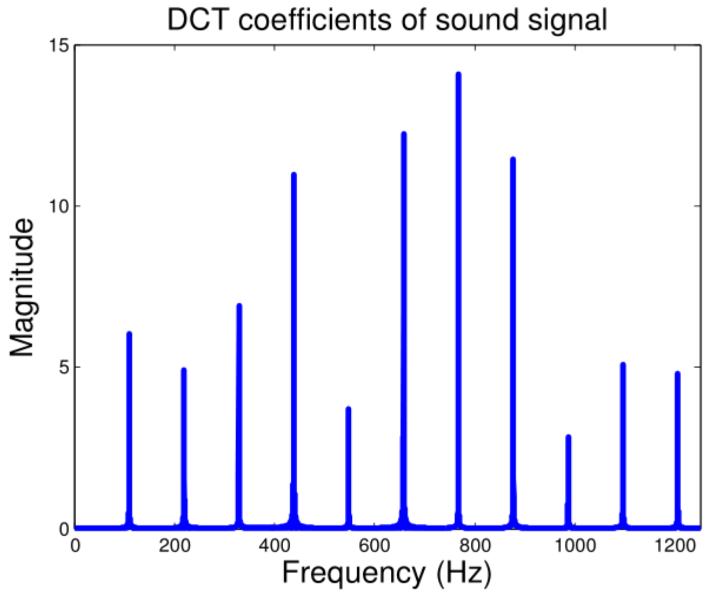
Example: Fourier Decomposition / Laplacian eigenfunctions

$$J(u) = \sqrt{\int |\nabla u|^2 \; dx}$$

Laplacian Eigenfunctions

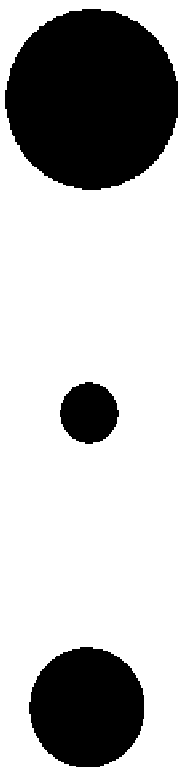
Fourier cosine decomposition in 1D





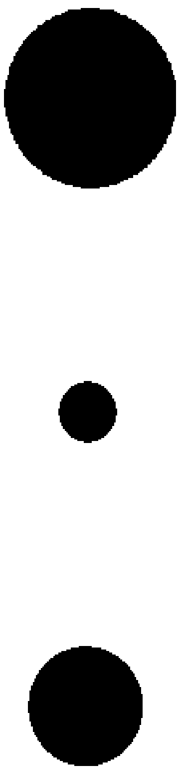
Spectral Decomposition

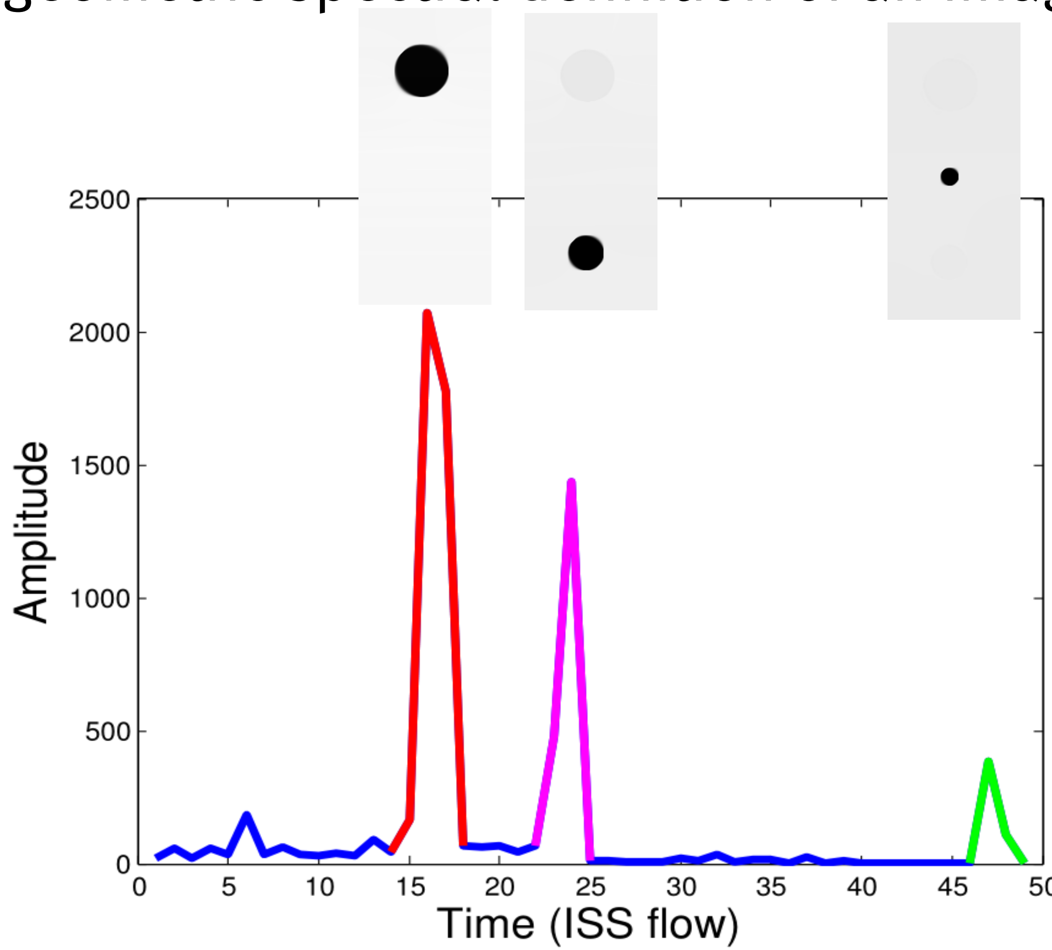
Natural / geometric spectral definition of an image ?



Spectral Decomposition

Natural / geometric spectral definition of an image ?





What is a Spectral Decomposition ?

Standard case related to positive semifinite linear operator *A* in Hilbert space *X*, respectively seminorm

$$J(u) = \sqrt{\langle Au, u \rangle}$$

Spectral theorem: there exists a vector valued measure (to the space of linear operators on *X*) such that for a scalar function

$$\varphi(A) = \int_{\mathbb{R}_+} \varphi(\lambda) \; dE_\lambda$$

In particular decomposition of *A* and Identity

Filtering

We are not interested in the operator, but in action on *f*

$$\varphi(A) f = \int_{\mathbb{R}_+} \varphi(\lambda) \; dE_\lambda \cdot f$$

Hence we have a vector valued measure into *X*

$$\tilde{\Phi}_s = E_s \cdot f$$

Spectral decomposition / filtering

$$\varphi(A) f = \int_{\mathbb{R}_+} \varphi(t) \; d\tilde{\Phi}_s$$

Nonlinear Spectral Decomposition

Keep basic properties:

Definition 1. *A map from $f \in \mathcal{X}$ to a vector-valued Radon measure $\tilde{\Phi}_s$ on $\mathcal{X}$ is called a* spectral (frequency) representation *with respect to the convex functional J if the following properties are satisfied:*

- **Eigenvectors as atoms:** *For f satisfying $\|f\| = 1$ and $\lambda f \in \partial J(f)$ the spectral representation is given by $\tilde{\Phi}_s = f\ \delta_\lambda(s)$.*
- **Reconstruction:** *The input data f, for any $f \in \mathcal{X}$, can be reconstructed by*

$$f = \int_0^\infty d\tilde{\Phi}_s.$$

Wavelength representation: $$d\Phi_t = \frac{1}{s^2} \tilde{d}\Phi_{1/s}$$

Nonlinear Spectral Decomposition

Polar decomposition of the measure defines spectrum

$$d\tilde{\Phi}_s = \tilde{\psi}(s)\ d\|\tilde{\Phi}_s\|$$

Parseval identity

$$\|f\|^2 = \int_0^\infty d(\tilde{\Phi}_s \cdot f) = \int_0^\infty (\tilde{\psi}(s) \cdot f) d\|\tilde{\Phi}_s\|$$

Defining Spectral Decompositions

Different options: variational methods / gradient flows

$$u_{VM}(t) = \arg\min_u \frac{1}{2}\|u - f\|_2^2 + tJ(u)$$

$$\partial_t u_{GF}(t) = -p_{GF}(t), \qquad p_{GF}(t) \in \partial J(u_{GF}(t)), \; u(0) = f$$

$$\partial_s q_{IS}(s) = f - v_{IS}(s), \qquad q_{IS}(s) \in \partial J(v_{IS}(s)), \; q_{IS}(0) = 0.$$

Defining Spectral Decompositions

Spectral representation derived from dynamics of eigenfunctions

$$\phi_{VM}(t) = t\partial_{tt} u_{VM}(t)$$

$$\phi_{GF}(t) = t\partial_{tt} u_{GF}(t)$$

$$\tilde{\phi}_{IS}(s) = \partial_s v_{IS}(s) = -\partial_{ss} q_{IS}(s)$$

$$\phi_{IS}(t) = -\partial_s v_{IS}(\frac{1}{t}) = t^2 \partial_t u_{IS}(t)$$

Connections of Spectral Decompositions

Very similar results for all the spectral decompositions

Conjecture: under appropriate conditions all spectral representations are the same

GF and VM have same primal variable *u(t)*

VM and IS have same dual variable *p(t) / q(s)*

VM dual variable is Fejer mean of GF dual variable
VM primal variable is Fejer mean of IS primal variable

Relations of Spectral Decomposition

Consider finite-dimensional polyhedral (crystalline) case

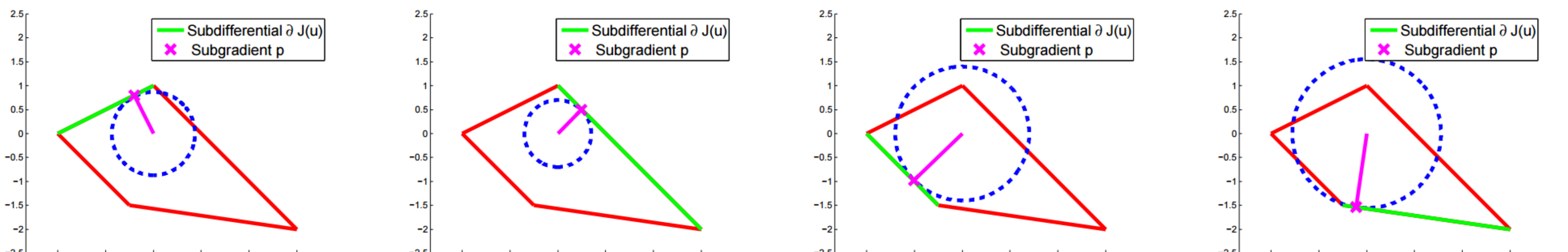
Definition (Polyhedral Seminorm (PS)). *We say that J induces a polyhedral seminorm (PS) if there exists a finite set whose convex hull equals $\partial J(0)$.*

Definition (MINSUB). *We say that J meets (MINSUB) if for all $u \in R^n$, the element $\hat{p}$ determined by*

$$\hat{p} = \arg\min_{p} \|p\|^2 \text{ subject to } p \in \partial J(u),$$

meets

$$\langle \hat{p}, \hat{p} - q \rangle = 0 \quad \forall q \in \partial J(u).$$



Piecewise linear dynamics under (PS)

Gradient flow (related to results by *Briani et 2011* for TV flow):

$$u_{GF}(t) = u_{GF}(t_i) - (t - t_i) p_{GF}(t_{i+1})$$

$$p_{GF}(t_{i+1}) \ \in \ \partial J(u_{GF}(t)) \ \ for \ t \ \in \ [t_i, t_{i+1}]$$

Inverse scale space method (related to *mb-Möller-Benning-Osher 2012, Möller-mb 2014*):

$$q_{IS}(s) = q_{IS}(s_i) - (s - s_i)(f - v_{IS}(s_{i+1}))$$

$$q_{IS}(s) \ \in \ \partial J(v_{IS}(s_{i+1})) \ \ for \ s \ \in \ [s_i, s_{i+1}]$$

Piecewise linear dynamics under (PS)

Variational method:

$$u_{VM} \ \ is \ \ an \ \ affinely \ \ linear \ \ function \ \ of \ t \ \ for \ t \in [t_i, t_{i+1}]$$

$$q_{VM}(s) = p_{VM}(1/s) \ \ is \ \ an \ \ affinely \ \ linear \ \ function \ \ of \ \ s \ \ for \ \ s \in [s_i, s_{i+1}], \ s_i = \frac{1}{t_i}$$

$$\phi_*(t) = \sum_{i=0}^{N_*} \phi_*^i \delta(t - t_i), \qquad for \ * \in \{VM, GF, IS\}$$

Spectral representation

Well-defined decomposition

$$\phi_*(t) = \sum_{i=0}^{N_*} \phi_*^i \delta(t - t_i), \qquad for * \in \{VM, GF, IS\}$$

$$f = \sum_{i=0}^{N_*} \phi_*^i, \qquad for * \in \{VM, GF, IS\}$$

Equivalence under (MINSUB)

If (PS) and (MINSUB) hold, then for all *f*:

$$u_{GF}(t) = u_{VM}(t)$$

$$p_{VM}(t) = \frac{1}{t}\int_0^t p_{GF}(s)ds \in \partial J(u_*(t))$$

Equivalence under (DD-L1)

Canonical example $J(u) = \|Ku\|_1$

(MINSUB) is satisfied if and only if *KK** is weakly diagonally dominant

Satisfied e.g. for 1D TV, K = div *(Briani et al 2011)*

Under this condition we also obtain that VM and IS are equivalent (same dual variable)

In particular all three approaches yield the same spectral decomposition

Eigenfunction decomposition

Under (DD-L1) the subgradients of the gradient flow are eigenfunctions of J

Hence we have a decomposition into eigenfunctions

$$f = \sum_{j=0}^{N} (t_{j+1} - t_j) p_{GF}(t_{j+1})$$

$$\partial_t u_{GF}(t) = -p_{GF}(t), \qquad p_{GF}(t) \in \partial J(u_{GF}(t)), \; u(0) = f$$

Applications: Filtering

Applications: Filtering

Applications: Ageing

Applications: Ageing

Applications: Personalized Avatar

Conclusion

Ground states and singular vectors can be generalized nicely to nonlinear setup

Yield detailed insight into behaviour of regularization methods and multiple scales

Potential for further investigation

Computation of singular vectors (explicit / numerical)