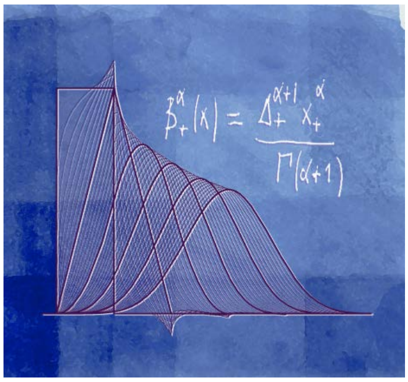
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

# The rise and fall of regularizations: $l_1$ vs. $l_2$ representer theorems

Michael Unser
Biomedical Imaging Group
EPFL, Lausanne, Switzerland

Joint work with
Julien Fageot,
John-Paul Ward, and Harshit Gupta



London Workshop on Sparse Signal Processing, 15-16 Sept. 2016, Imperial College.

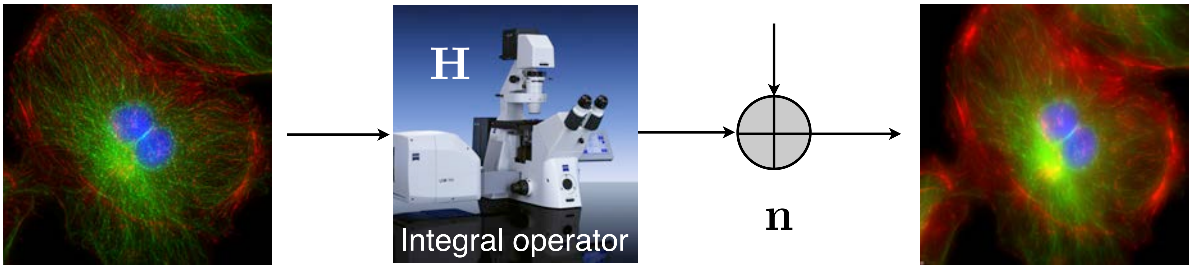
## OUTLINE

- **Linear inverse problems and regularization**
  - Tikhonov regularization
  - The sparsity (r)evolution
  - Compressed sensing and $l_1$ minimization
- **Part I: Discrete-domain regularization ($l_2$ vs. $l_1$)**
- **Part II: Continuous-domain regularization ($L_2$ vs. gTV)**
  - Classical $L_2$ regularization: theory of RKHS
  - Splines and operators
  - Minimization of gTV: the optimality of splines
  - Enabling components for the proof
  - Special case TV in 1D

FNSNF SWISS NATIONAL SCIENCE FOUNDATION
erc

# Inverse problems in bio-imaging

- Linear forward model $\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}$




$\mathbf{s}$

Problem: recover $\mathbf{s}$ from noisy measurements $\mathbf{y}$

- The easy scenario

Inverse problem is **well**

$\Rightarrow \quad \mathbf{s} \approx \mathbf{H}^{-1}\mathbf{y}$

- Backprojection (po

## Basic limitations

1) Inherent noise amplification
2) Difficulty to invert **H** (too large or non-square)
3) All interesting inverse problems are **ill-posed**

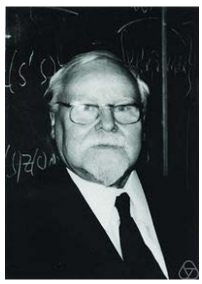
# Linear inverse problems (20th century theory)

- Dealing with **ill-posed problems**: Tikhonov **regularization**

$\mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_2^2$: regularization (or smoothness) functional

$\mathbf{L}$: regularization operator (i.e., Gradient)

$$\min_{\mathbf{s}} \mathcal{R}(\mathbf{s}) \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2 \leq \sigma^2$$



*Andrey N. Tikhonov* (1906-1993)

- Equivalent variational problem

$$\mathbf{s}^\star = \arg\min \underbrace{\|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2}_{\text{data consistency}} + \underbrace{\lambda\|\mathbf{L}\mathbf{s}\|_2^2}_{\text{regularization}}$$

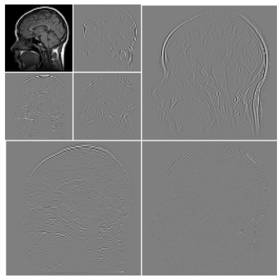
Formal linear solution: $\mathbf{s} = (\mathbf{H}^T\mathbf{H} + \lambda\mathbf{L}^T\mathbf{L})^{-1}\mathbf{H}^T\mathbf{y} = \mathbf{R}_\lambda \cdot \mathbf{y}$

Interpretation: "filtered" backprojection

# Linear inverse problems: The sparsity (r)evolution

(20th Century) $p = 2 \longrightarrow 1$ (21st Century)

$$\mathbf{s}_{\text{rec}} = \arg\min_{\mathbf{s}} \left( \|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2 + \lambda \mathcal{R}(\mathbf{s}) \right)$$



- Non-quadratic regularization regularization

  $\mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_{\ell_2}^2 \longrightarrow \|\mathbf{L}\mathbf{s}\|_{\ell_p}^p \longrightarrow \|\mathbf{L}\mathbf{s}\|_{\ell_1}$

  - Total variation (Rudin-Osher, 1992)

    $\mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_{\ell_1}$ with $\mathbf{L}$: gradient

  - Wavelet-domain regularization (Figuereido et al., Daubechies et al. 2004)

    $\mathbf{v} = \mathbf{W}^{-1}\mathbf{s}$: wavelet expansion of s (typically, sparse)

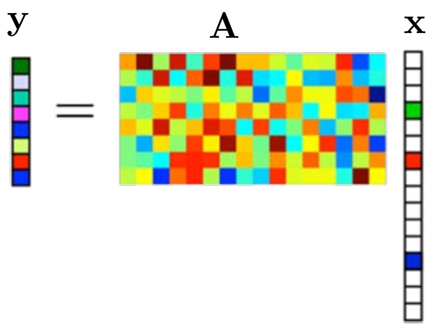
    $\mathcal{R}(\mathbf{s}) = \|\mathbf{v}\|_{\ell_1}$

- Compressed sensing/sampling (Candes-Romberg-Tao; Donoho, 2006)



# Compressive sensing (CS) and $l_1$ minimization

[Donoho et al., 2005
Candès-Tao, 2006, ...]




+ “noise”

Sparse representation of signal: $\mathbf{s} = \mathbf{W}\mathbf{x}$ with $\|\mathbf{x}\|_0 = K \ll N_x$

Equivalent $N_y \times N_x$ **sensing matrix** : $\mathbf{A} = \mathbf{H}\mathbf{W}$

- Constrained (synthesis) formulation of recovery problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \le \sigma^2$$

## CS: Three fundamental ingredients

(Donoho, *IEEE T. Inf. Theo.* 2006) (Candès-Romberg, *Inv. Prob.* 2007)

1. Existence of sparsifying transform (**W** or **L**)
   - Wavelet basis
   - Dictionary
   - Differential operator (Gradient)

2. Incoherence of sensing matrix **A**
   - Restricted isometry; few linearly dependent columns (spark)
   - Quasi-random and delocalized structure:
     *Gaussian matrix with i.i.d. entries,*
     *random sampling in Fourier domain*

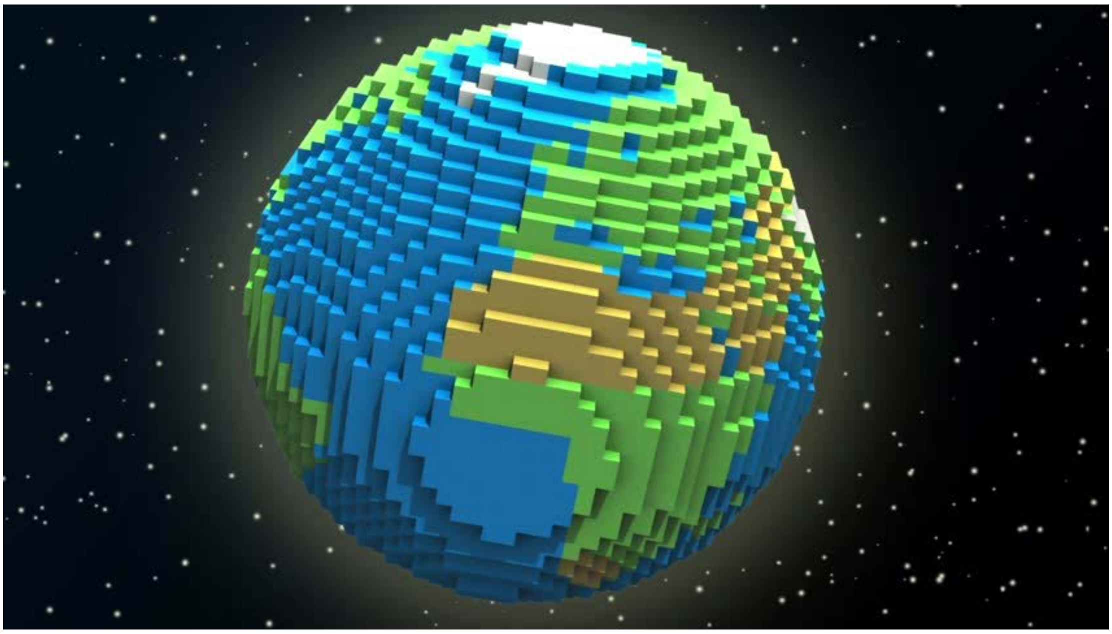
3. Non-linear signal recovery ($l_1$ minimization)



## CS: Examples of applications in imaging

- Magnetic resonance imaging (MRI) (Lustig, *Mag. Res. Im.* 2007)
- Radio Interferometry (Wiaux, *Notic. R. Astro.* 2007)
- Teraherz Imaging (Chan, *Appl. Phys.* 2008)
- Digital holography (Brady, *Opt. Express* 2009; Marim 2010)
- Spectral-domain OCT (Liu, *Opt. Express* 2010)
- Coded-aperture spectral imaging (Arce, *IEEE Sig. Proc.* 2014)
- Localization microscopy (Zhu, *Nat. Meth.* 2012)
- Ultrafast photography (Gao, *Nature* 2014)

# Part I: Discrete-domain regularization





# Classical regularized least-squares estimator

- Linear measurement model:
  $y_m = \langle \mathbf{h}_m, \mathbf{x} \rangle + n[m], \quad m = 1, \dots, M$
- System matrix of size $M \times N$ : $\mathbf{H} = [\mathbf{h}_1 \cdots \mathbf{h}_M]^T$

$$\mathbf{x}_{\mathrm{LS}} = \arg\min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_2^2$$

$$\Rightarrow \quad \mathbf{x}_{\mathrm{LS}} = (\mathbf{H}^T\mathbf{H} + \lambda \mathbf{I}_N)^{-1}\mathbf{H}^T\mathbf{y}$$

$$= \mathbf{H}^T\mathbf{a} = \sum_{m=1}^{M} a_m \mathbf{h}_m \quad \text{where} \quad \mathbf{a} = (\mathbf{H}\mathbf{H}^T + \lambda \mathbf{I}_M)^{-1}\mathbf{y}$$

Interpretation: $\mathbf{x}_{\mathrm{LS}} \in \mathrm{span}\{\mathbf{h}_m\}_{m=1}^{M}$

Lemma
$(\mathbf{H}^T\mathbf{H} + \lambda \mathbf{I}_N)^{-1}\mathbf{H}^T = \mathbf{H}^T(\mathbf{H}\mathbf{H}^T + \lambda \mathbf{I}_M)^{-1}$

# Generalization: constrained $l_2$ minimization

- Discrete signal to reconstruct: $x = (x[n])_{n\in\mathbb{Z}}$
- Sensing operator $\mathrm{H} : \ell_2(\mathbb{Z}) \to \mathbb{R}^M$
  $x \mapsto \mathbf{z} = \mathrm{H}\{x\} = (\langle x, h_1\rangle, \dots, \langle x, h_M\rangle)$ with $h_m \in \ell_2(\mathbb{Z})$
- Closed convex set in measurement space: $\mathcal{C} \subset \mathbb{R}^M$

Example: $\mathcal{C}_{\mathbf{y}} = \{\mathbf{z} \in \mathbb{R}^M : \|\mathbf{y} - \mathbf{z}\|_2^2 \le \sigma^2\}$

**Representer theorem for constrained $\ell_2$ minimization**

(P2) $\min_{x\in\ell_2(\mathbb{Z})} \|x\|_{\ell_2}^2$ s.t. $\mathrm{H}\{x\} \in \mathcal{C}$

The problem (P2) has a unique solution of the form

$$x_{\mathrm{LS}} = \sum_{m=1}^{M} a_m h_m = \mathrm{H}^*\{\mathbf{a}\}$$

with expansion coefficients $\mathbf{a} = (a_1, \cdots, a_M) \in \mathbb{R}^M$.

(U.-Fageot-Gupta *IEEE Trans. Info. Theory*, Sept. 2016)



# Constrained $l_1$ minimization $\Rightarrow$ sparsifying effect

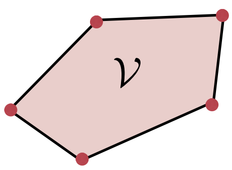
- Discrete signal to reconstruct: $x = (x[n])_{n\in\mathbb{Z}}$
- Sensing operator $\mathrm{H} : \ell_1(\mathbb{Z}) \to \mathbb{R}^M$
  $x \mapsto \mathbf{z} = \mathrm{H}\{x\} = (\langle x, h_1\rangle, \dots, \langle x, h_M\rangle)$ with $h_m \in \ell_\infty(\mathbb{Z})$
- Closed convex set in measurement space: $\mathcal{C} \subset \mathbb{R}^M$

**Representer theorem for constrained $\ell_1$ minimization**

(P1) $\mathcal{V} = \arg\min_{x\in\ell_1(\mathbb{Z})} \|x\|_{\ell_1}$ s.t. $\mathrm{H}\{x\} \in \mathcal{C}$

is convex, weak*-compact with extreme points of the form

$$x_{\mathrm{sparse}}[\cdot] = \sum_{k=1}^{K} a_k \delta[\cdot - n_k] \quad \text{with} \quad K = \|x_{\mathrm{sparse}}\|_0 \le M.$$



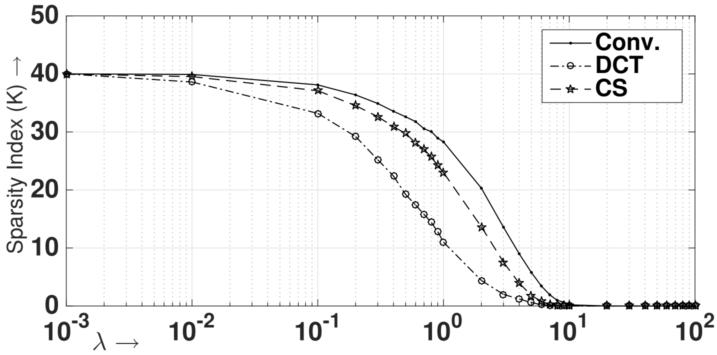

If CS condition is satisfied, then solution is unique

(U.-Fageot-Gupta *IEEE Trans. Info. Theory*, Sept. 2016)

# Controlling sparsity

Measurement model: $y_m = \langle h_m, x \rangle + n[m], \quad m = 1, \ldots, M$

$$x_{\text{sparse}} = \arg \min_{x \in \ell_1(\mathbb{Z})} \left( \sum_{m=1}^{M} \left| y_m - \langle h_m, x \rangle \right|^2 + \lambda \|x\|_{\ell_1} \right)$$



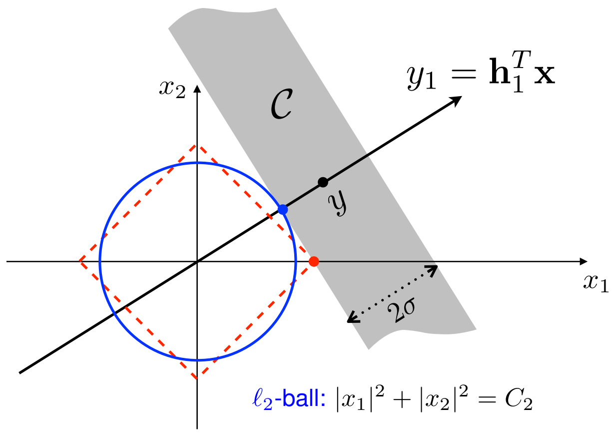



# Geometry of $l_2$ vs. $l_1$ minimization

- Prototypical inverse problem

$$\min_{\mathbf{x}} \left\{ \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 + \lambda \|\mathbf{x}\|_{\ell_2}^2 \right\} \quad \Leftrightarrow \quad \min_{\mathbf{x}} \|\mathbf{x}\|_{\ell_2} \text{ subject to } \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 \le \sigma^2$$

$$\min_{\mathbf{x}} \left\{ \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 + \lambda \|\mathbf{x}\|_{\ell_1} \right\} \quad \Leftrightarrow \quad \min_{\mathbf{x}} \|\mathbf{x}\|_{\ell_1} \text{ subject to } \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 \le \sigma^2$$




$\ell_2$-ball: $|x_1|^2 + |x_2|^2 = C_2$

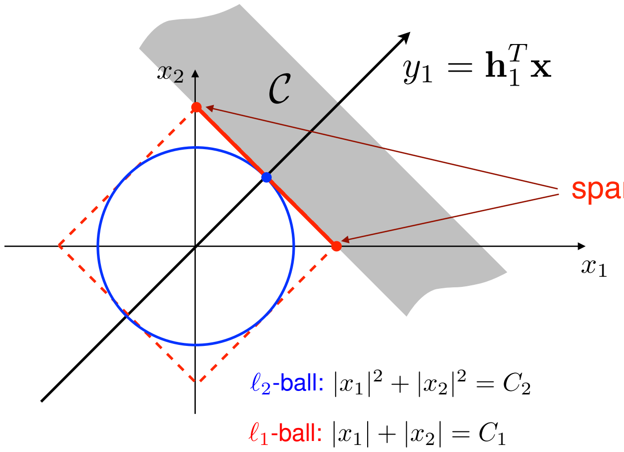
$\ell_1$-ball: $|x_1| + |x_2| = C_1$

# Geometry of $l_2$ vs. $l_1$ minimization

- Prototypical inverse problem

$$\min_{\mathbf{x}} \left\{ \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 + \lambda \|\mathbf{x}\|_{\ell_2}^2 \right\} \Leftrightarrow \min_{\mathbf{x}} \|\mathbf{x}\|_{\ell_2} \text{ subject to } \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 \le \sigma^2$$

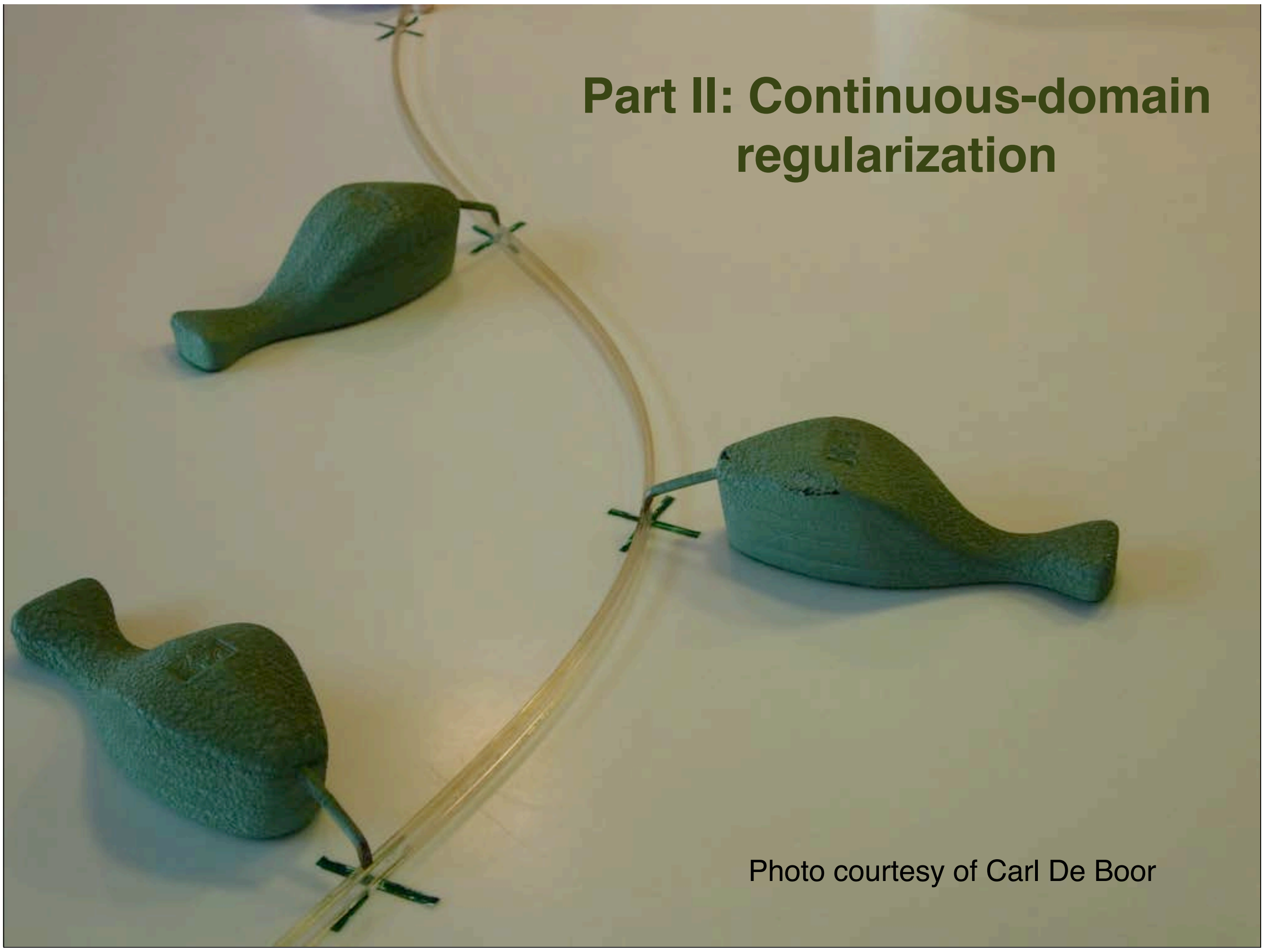
$$\min_{\mathbf{x}} \left\{ \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 + \lambda \|\mathbf{x}\|_{\ell_1} \right\} \Leftrightarrow \min_{\mathbf{x}} \|\mathbf{x}\|_{\ell_1} \text{ subject to } \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 \le \sigma^2$$




Configuration for **non-unique** $\ell_1$ solution



# Part II: Continuous-domain regularization




## Continuous-domain regularization ($L_2$ scenario)

Regularization functional: $\|\mathrm{L}f\|_{L_2}^2 = \int_{\mathbb{R}^d} |\mathrm{L}f(\boldsymbol{x})|^2 \mathrm{d}\boldsymbol{x}$

$\mathrm{L}$: suitable differential operator

- Theory of reproducing kernel Hilbert spaces (Aronszajn 1950)

  $\langle f, g \rangle_{\mathcal{H}} = \langle \mathrm{L}f, \mathrm{L}g \rangle$

- Interpolation and approximation theory
  - Smoothing splines (Schoenberg 1964, Kimeldorf-Wahba 1971)
  - Thin-plate splines, radial basis functions (Duchon 1977)
- Machine learning
  - Radial basis functions, kernel methods (Poggio-Girosi 1990)
  - Representer theorem(s) (Schölkopf-Smola 2001)

# Representer theorem for $L_2$ regularization

(P2) $$\arg\min_{f\in\mathcal{H}} \left( \sum_{m=1}^{M} |y_m - f(\boldsymbol{x}_m)|^2 + \lambda \|f\|_{\mathcal{H}}^2 \right)$$

$h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is the (unique) **reproducing kernel** for the Hilbert space $\mathcal{H}(\mathbb{R}^d)$ if
(i) $h(\boldsymbol{x}_0, \cdot) \in \mathcal{H}$ for all $\boldsymbol{x}_0 \in \mathbb{R}^d$
(ii) $f(\boldsymbol{x}_0) = \langle h(\boldsymbol{x}_0, \cdot), f\rangle_{\mathcal{H}}$ for all $f \in \mathcal{H}$ and $\boldsymbol{x}_0 \in \mathbb{R}^d$

Convex loss function: $F : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$

Sample values: $\boldsymbol{f} = \big(f(\boldsymbol{x}_1), \dots, f(\boldsymbol{x}_M)\big)$

(P2') $$\arg\min_{f\in\mathcal{H}} \left( F(\boldsymbol{y}, \boldsymbol{f}) + \lambda \|f\|_{\mathcal{H}}^2 \right)$$

(Schölkopf-Smola 2001)

**Representer theorem for $L_2$-regularization**

The generic parametric form of the solution of (P2') is

$$f(\boldsymbol{x}) = \sum_{m=1}^{M} a_m h(\boldsymbol{x}, \boldsymbol{x}_m)$$

Supports the theory of SVM, kernel methods, etc.



# Sparsity and continuous-domain modeling

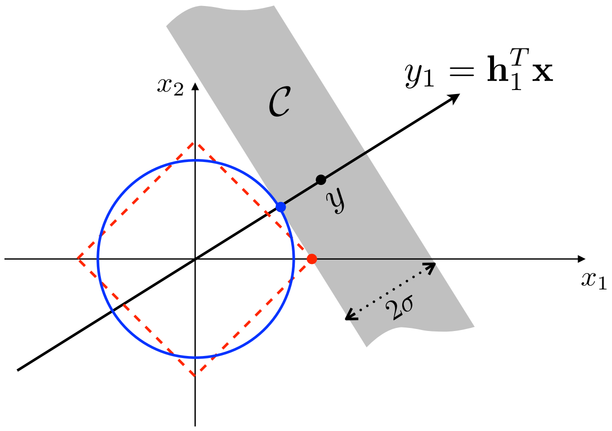
- Compressed sensing (CS)
  - Generalized sampling and infinite-dimensional CS (Adcock-Hansen, 2011)
  - Xampling: CS of analog signals (Eldar, 2011)
- Splines and approximation theory
  - $L_1$ splines (Fisher-Jerome, 1975)
  - Locally-adaptive regression splines (Mammen-van de Geer, 1997)
  - Generalized TV (Steidl et al. 2005; Bredies et al. 2010)
- Statistical modeling
  - Sparse stochastic processes (Unser et al. 2011-2014)

# Geometry of $l_2$ vs. $l_1$ minimization

- Prototypical inverse problem

$$\min_{\mathbf{x}} \left\{ \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 + \lambda \|\mathbf{x}\|_{\ell_2}^2 \right\} \Leftrightarrow \min_{\mathbf{x}} \|\mathbf{x}\|_{\ell_2} \text{ subject to } \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 \leq \sigma^2$$

$$\min_{\mathbf{x}} \left\{ \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 + \lambda \|\mathbf{x}\|_{\ell_1} \right\} \Leftrightarrow \min_{\mathbf{x}} \|\mathbf{x}\|_{\ell_1} \text{ subject to } \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 \leq \sigma^2$$



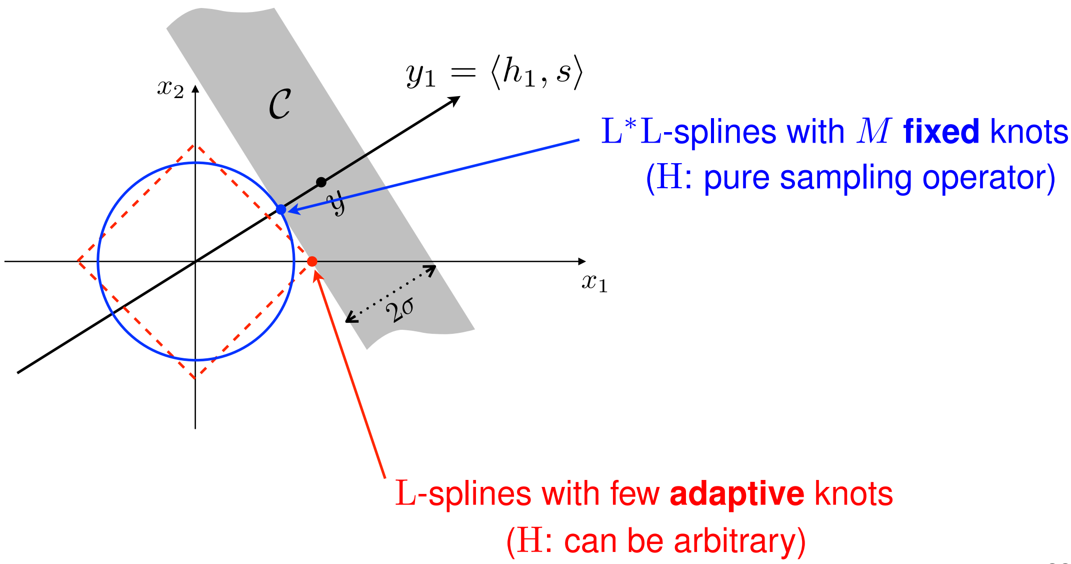



# "Geometry" of $L_2$ vs. TV minimization

- Prototypical inverse problem

$$\min_{s} \left\{ \|\mathbf{y} - \mathrm{H}\{s\}\|_{\ell_2}^2 + \lambda \|\mathrm{L}\{s\}\|_{L_2}^2 \right\} \Leftrightarrow \min_{s} \|\mathrm{L}\{s\}\|_{L_2}^2 \text{ subject to } \|\mathbf{y} - \mathrm{H}\{s\}\|_{\ell_2}^2 \leq \sigma^2$$

$$\min_{s} \left\{ \|\mathbf{y} - \mathrm{H}\{s\}\|_{\ell_2}^2 + \lambda \|\mathrm{L}\{s\}\|_{\mathrm{TV}} \right\} \Leftrightarrow \min_{s} \|\mathrm{L}\{s\}\|_{\mathrm{TV}} \text{ subject to } \|\mathbf{y} - \mathrm{H}\{s\}\|_{\ell_2}^2 \leq \sigma^2$$



# Splines are analog and intrinsically sparse

$\mathrm{L}\{\cdot\}$: admissible differential operator

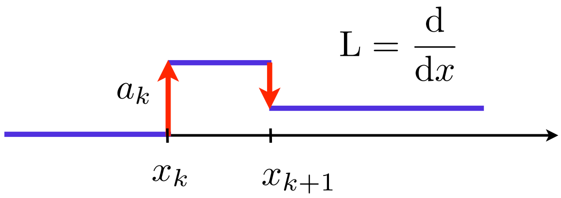
$\delta(\cdot - \boldsymbol{x}_0)$: Dirac impulse shifted by $\boldsymbol{x}_0 \in \mathbb{R}^d$

**Definition**

The function $s : \mathbb{R}^d \to \mathbb{R}$ is a (non-uniform) $\mathrm{L}$-spline with knots $(\boldsymbol{x}_k)_{k=1}^K$ if

$$\mathrm{L}\{s\} = \sum_{k=1}^{K} a_k \delta(\cdot - \boldsymbol{x}_k) = w_\delta \quad : \text{ spline's innovation}$$

Spline theory: (Schultz-Varga, 1967)




- FRI signal processing: Innovation variables $(2K)$ (Vetterli et al., 2002)
  - Location of singularities (knots) : $\{x_k\}_{k=1}^K$
  - Strength of singularities (linear weights): $\{a_k\}_{k=1}^K$

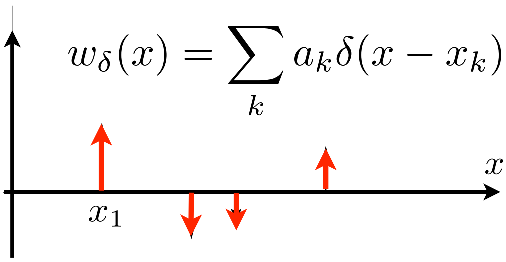


# Spline synthesis: example

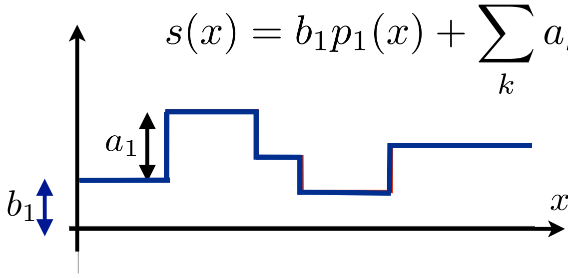
$\mathrm{L} = \mathrm{D} = \frac{\mathrm{d}}{\mathrm{d}x}$ Null space: $\mathcal{N}_\mathrm{D} = \mathrm{span}\{p_1\}, \quad p_1(x) = 1$

$\rho_\mathrm{D}(x) = \mathrm{D}^{-1}\{\delta\}(x) = \mathbb{1}_+(x)$: Heaviside function

$$w_\delta(x) = \sum_k a_k \delta(x - x_k)$$




$$s(x) = b_1 p_1(x) + \sum_k a_k \mathbb{1}_+(x - x_k)$$



# Spline synthesis: generalization

$\mathrm{L}$: spline admissible operator (LSI)

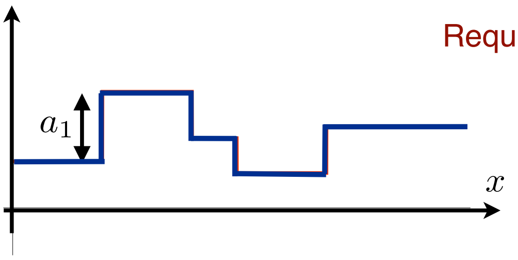
$\rho_{\mathrm{L}}(\boldsymbol{x}) = \mathrm{L}^{-1}\{\delta\}$: Green's function of $\mathrm{L}$

Finite-dimensional null space: $\mathcal{N}_{\mathrm{L}} = \mathrm{span}\{p_n\}_{n=1}^{N_0}$

Spline's innovation: $w_\delta(\boldsymbol{x}) = \sum_k a_k \delta(\boldsymbol{x} - \boldsymbol{x}_k)$

$$\Rightarrow \quad s(\boldsymbol{x}) = \sum_k a_k \rho_{\mathrm{L}}(\boldsymbol{x} - \boldsymbol{x}_k) + \sum_{n=1}^{N_0} b_n p_n(\boldsymbol{x})$$

Requires specification of boundary conditions






# Principled operator-based approach

- Biorthogonal basis of $\mathcal{N}_{\mathrm{L}} = \mathrm{span}\{p_n\}_{n=1}^{N_0}$
  - $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_{N_0})$ such that $\langle \phi_m, p_n \rangle = \delta_{m,n}$
  - Projection operator: $p = \sum_{n=1}^{N_0} \langle \phi_n, p \rangle p_n$ for all $p \in \mathcal{N}_{\mathrm{L}}$
- Operator-based spline synthesis
  - Boundary conditions: $\langle s, \phi_n \rangle = b_n, \; n = 1, \cdots, N_0$
  - Spline's innovation: $\mathrm{L}\{s\} = w_\delta = \sum_k a_k \delta(\cdot - \boldsymbol{x}_k)$

$$s(\boldsymbol{x}) = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{w_\delta\}(\boldsymbol{x}) + \sum_{n=1}^{N_0} b_n p_n(\boldsymbol{x})$$

- Existence of $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ as a stable right-inverse of $\mathrm{L}$ ? (see **Theorem 1**)
  - $\mathrm{L}\mathrm{L}_{\boldsymbol{\phi}}^{-1} w = w$
  - $\boldsymbol{\phi}(\mathrm{L}_{\boldsymbol{\phi}}^{-1} w) = \mathbf{0}$


# From Dirac impulses to Borel measures

$\mathcal{S}(\mathbb{R}^d)$: Schwartz's space of smooth and rapidly decaying test functions on $\mathbb{R}^d$

$\mathcal{S}'(\mathbb{R}^d)$: Schwartz's space of tempered distributions

- Space of real-valued, countably additive Borel measures on $\mathbb{R}^d$

$$\mathcal{M}(\mathbb{R}^d) = \big(C_0(\mathbb{R}^d)\big)' = \big\{w \in \mathcal{S}'(\mathbb{R}^d) : \|w\|_{\mathcal{M}} = \sup_{\varphi \in \mathcal{S}(\mathbb{R}^d):\|\varphi\|_\infty = 1} \langle w, \varphi\rangle < \infty\big\},$$

where $w : \varphi \mapsto \langle w, \varphi\rangle = \int_{\mathbb{R}^d} \varphi(\boldsymbol{r}) w(\boldsymbol{r}) \mathrm{d}\boldsymbol{r}$

- Equivalent definition of "total variation" norm

$$\|w\|_{\mathcal{M}} = \sup_{\varphi \in C_0(\mathbb{R}^d):\|\varphi\|_\infty = 1} \langle w, \varphi\rangle$$

- Basic inclusions
  - $\delta(\cdot - \boldsymbol{x}_0) \in \mathcal{M}(\mathbb{R}^d)$ with $\|\delta(\cdot - \boldsymbol{x}_0)\|_{\mathcal{M}} = 1$ for any $\boldsymbol{x}_0 \in \mathbb{R}^d$
  - $\|f\|_{\mathcal{M}} = \|f\|_{L_1(\mathbb{R}^d)}$ for all $f \in L_1(\mathbb{R}^d)$ $\quad\Rightarrow\quad$ $L_1(\mathbb{R}^d) \subseteq \mathcal{M}(\mathbb{R}^d)$

# Optimality result for Dirac measures

- $\mathbf{F}$: linear continuous map $\mathcal{M}(\mathbb{R}^d) \to \mathbb{R}^M$
- $\mathcal{C}$: convex compact subset of $\mathbb{R}^M$
- Generic constrained TV minimization problem

$$\mathcal{V} = \arg \min_{w \in \mathcal{M}(\mathbb{R}^d)\,:\,\mathbf{F}(w) \in \mathcal{C}} \|w\|_{\mathcal{M}}$$

**Generalized Fisher-Jerome theorem**

The solution set $\mathcal{V}$ is a **convex, weak$*$-compact** subset of $\mathcal{M}(\mathbb{R}^d)$ with **extremal points** of the form

$$w_\delta = \sum_{k=1}^{K} a_k \delta(\cdot - \boldsymbol{x}_k)$$

with $K \leq M$ and $\boldsymbol{x}_k \in \mathbb{R}^d$. (U.-Fageot-Ward, ArXiv 2016)

Jerome-Fisher, 1975: Compact domain & scalar intervals



# General convex problems with gTV regularization

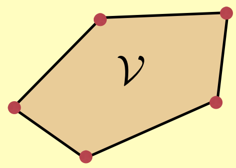
$$\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) = \left\{ s : \mathrm{gTV}(s) = \|\mathrm{L}\{s\}\|_{\mathcal{M}} = \sup_{\|\varphi\|_\infty \leq 1} \langle \mathrm{L}\{s\}, \varphi \rangle < \infty \right\}$$

- **Linear** measurement operator $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \to \mathbb{R}^M : f \mapsto \mathbf{z} = \mathrm{H}\{f\}$
- $\mathcal{C}$: **convex** compact subset of $\mathbb{R}^M$
- Finite-dimensional **null space** $\mathcal{N}_{\mathrm{L}} = \{q \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) : \mathrm{L}\{q\} = 0\}$ with basis $\{p_n\}_{n=1}^{N_0}$

Admissibility of regularization: $\mathrm{H}\{q_1\} = \mathrm{H}\{q_2\} \Leftrightarrow q_1 = q_2$ for all $q_1, q_2 \in \mathcal{N}_{\mathrm{L}}$

**Representer theorem for gTV regularization**

The extremal points of the constrained minimization problem




$$\mathcal{V} = \arg \min_{f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)} \|\mathrm{L}\{f\}\|_{\mathcal{M}} \quad \text{s.t.} \quad \mathrm{H}\{f\} \in \mathcal{C}$$

are necessarily of the form $f(\boldsymbol{x}) = \sum_{k=1}^{K} a_k \rho_{\mathrm{L}}(\boldsymbol{x} - \boldsymbol{x}_k) + \sum_{n=1}^{N_0} b_n p_n(\boldsymbol{x})$ with $K \leq M - N_0$; that is, **non-uniform** $\mathrm{L}$**-splines** with knots at the $\boldsymbol{x}_k$ and $\|\mathrm{L}\{f\}\|_{\mathcal{M}} = \sum_{k=1} |a_k|$. The full solution set is the **convex hull** of those extremal points.

(U.-Fageot-Ward, ArXiv 2016)

# Enabling components for proof of the theorem



# Existence of stable right-inverse operator

$$L_{\infty,n_0}(\mathbb{R}^d) = \{f : \mathbb{R}^d \to \mathbb{R} : \sup_{\boldsymbol{x}\in\mathbb{R}^d} \left(|f(\boldsymbol{x})|(1 + \|\boldsymbol{x}\|)^{-n_0}\right) < +\infty\}$$

**Theorem 1** (U.-Fageot-Ward, ArXiv 2016)
Let $\mathrm{L}$ be a spline-admissible operator with a $N_0$-dimensional null space $\mathcal{N}_{\mathrm{L}} \subseteq L_{\infty,n_0}(\mathbb{R}^d)$ such that $p = \sum_{n=1}^{N_0} \langle p, \phi_n \rangle p_n$ for all $p \in \mathcal{N}_{\mathrm{L}}$. Then, there exists a **unique and stable operator** $\mathrm{L}_{\boldsymbol{\phi}}^{-1} : \mathcal{M}(\mathbb{R}^d) \to L_{\infty,n_0}(\mathbb{R}^d)$ such that, for all $w \in \mathcal{M}(\mathbb{R}^d)$,

- Right-inverse property: $\mathrm{L}\mathrm{L}_{\boldsymbol{\phi}}^{-1} w = w$,
- Boundary conditions: $\boldsymbol{\phi}(\mathrm{L}_{\boldsymbol{\phi}}^{-1} w) = \mathbf{0}$ with $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_{N_0})$.

Its generalized impulse response $g_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{\delta(\cdot - \boldsymbol{y})\}(\boldsymbol{x})$ is given by

$$g_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{y}) = \rho_{\mathrm{L}}(\boldsymbol{x} - \boldsymbol{y}) - \sum_{n=1}^{N_0} p_n(\boldsymbol{x}) q_n(\boldsymbol{y})$$

with $\rho_{\mathrm{L}}$ such that $\mathrm{L}\{\rho_{\mathrm{L}}\} = \delta$ and $q_n(\boldsymbol{y}) = \langle \phi_n, \rho_{\mathrm{L}}(\cdot - \boldsymbol{y}) \rangle$.

# Characterization of generalized Beppo-Levi spaces

- Regularization operator $\mathrm{L}: \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \to \mathcal{M}(\mathbb{R}^d)$

$$f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \quad \Leftrightarrow \quad \mathrm{gTV}(f) = \|\mathrm{L}\{f\}\|_{\mathcal{M}} < \infty$$

**Theorem 2** (U.-Fageot-Ward, ArXiv 2016)

Let $\mathrm{L}$ be a spline-admissible operator that admits a stable right-inverse $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ of the form specified by Theorem 1. Then, any $f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ has a unique representation as

$$f = \mathrm{L}_{\boldsymbol{\phi}}^{-1} w + p,$$

where $w = \mathrm{L}\{f\} \in \mathcal{M}(\mathbb{R}^d)$ and $p = \sum_{n=1}^{N_0} \langle \phi_n, f \rangle p_n \in \mathcal{N}_{\mathrm{L}}$ with $\phi_n \in \big(\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)\big)'$. Moreover, $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ is a Banach space equipped with the norm

$$\|f\|_{\mathrm{L},\boldsymbol{\phi}} = \|\mathrm{L}f\|_{\mathcal{M}} + \|\boldsymbol{\phi}(f)\|_2.$$

- Generalized Beppo-Levi space: $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) = \mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d) \oplus \mathcal{N}_{\mathrm{L}}$

$$\mathcal{M}_{\mathrm{L},\boldsymbol{\phi}}(\mathbb{R}^d) = \{f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) : \boldsymbol{\phi}(f) = \mathbf{0}\}$$

$$\mathcal{N}_{\mathrm{L}} = \{p \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) : \mathrm{L}\{p\} = 0\}$$



# Example: Convex problem with TV regularization

$$\mathrm{L} = \mathrm{D} = \frac{\mathrm{d}}{\mathrm{d}x}$$

$$\mathcal{N}_{\mathrm{D}} = \mathrm{span}\{p_1\}, \quad p_1(x) = 1$$

$\rho_{\mathrm{D}}(x) = \mathbb{1}_+(x)$: Heaviside function

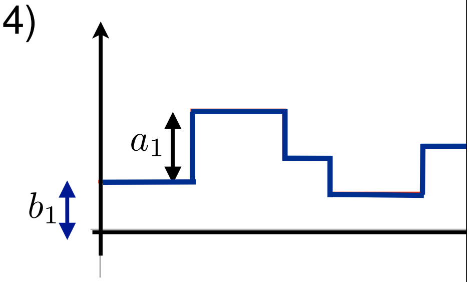
- General linear-inverse problem with TV regularization

$$\min_{s \in \mathcal{M}_{\mathrm{D}}(\mathbb{R})} \|\mathrm{D}\{s\}\|_{\mathcal{M}} \quad \text{s.t.} \quad \mathrm{H}\{s\} = (\langle h_1, s\rangle, \cdots, \langle h_M, s\rangle) \in \mathcal{C}(\mathbf{y})$$

- Generic form of the solution (by Theorem 4)

$$s(x) = b_1 + \sum_{k=1}^{K} a_k \mathbb{1}_+(x - x_k)$$

no penalty




with $K < M$ and free parameters $b_1$ and $(a_k, x_k)_{k=1}^{K}$

# SUMMARY: Sparsity in infinite dimensions

- Discrete-domain formulation
  - Contrasting behavior of $l_1$ vs. $l_2$ regularization
  - Minimization of $l_1$ favors sparse solutions (independently of sensing matrix)

- Continuous-domain formulation
  $s \in \mathcal{X}$
  - Linear measurement model $s \mapsto \mathbf{z} = \mathrm{H}\{s\}$
  - Linear signal model: PDE $\mathrm{L}s = w \quad \Rightarrow \quad s = \mathrm{L}^{-1}w$
  - L-splines = signals with "sparsest" innovation $\mathrm{gTV}(s) = \|\mathrm{L}s\|_{\mathcal{M}}$

- Deterministic optimality result
  - gTV **regularization**: favors "sparse" innovations
  - Non-uniform L-splines: ***universal*** solutions of linear inverse problems



# Acknowledgments

Many thanks to (former) members of EPFL's Biomedical Imaging Group

- Dr. Pouya Tafti
- Prof. Arash Amini
- Dr. Emrah Bostan
- Dr. Masih Nilchian
- Dr. Ulugbek Kamilov
- Dr. Cédric Vonesch
- ....


and collaborators ...

- Prof. Demetri Psaltis
- Prof. Marco Stampanoni
- Prof. Carlos-Oscar Sorzano
- Dr. Arne Seitz
- ....

erc 2

- Preprints and demos: http://bigwww.epfl.ch/

# References

- New results on sparsity-promoting regularization

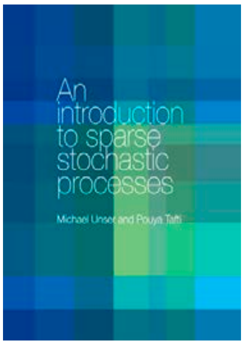

  - M. Unser, J. Fageot, H. Gupta, "Representer Theorems for Sparsity-Promoting $\ell_1$ Regularization," *IEEE Trans. Information Theory*, Vol. 62, No. 9, pp. 5167-5180.
  - M. Unser, J. Fageot, J.P. Ward, "Splines Are Universal Solutions of Linear Inverse Problems with Generalized-TV Regularization," arXiv:1603.01427 [math.FA].


- Theory of sparse stochastic processes


  - M. Unser and P. Tafti, ***An Introduction to Sparse Stochastic Processes***, Cambridge University Press, 2014.
    Preprint, available at `http://www.sparseprocesses.org`.

  - **For splines**: see chapter 6

CAMBRIDGE UNIVERSITY PRESS




- Algorithms and imaging applications


  - E. Bostan, U.S. Kamilov, M. Nilchian, M. Unser, "Sparse Stochastic Processes and Discretization of Linear Inverse Problems," *IEEE Trans. Image Processing*, vol. 22, no. 7, pp. 2699-2710, 2013.
  - C. Vonesch, M. Unser, "A Fast Multilevel Algorithm for Wavelet-Regularized Image Restoration," *IEEE Trans. Image Processing*, vol. 18, no. 3, pp. 509-523, March 2009.
  - M. Nilchian, C. Vonesch, S. Lefkimmiatis, P. Modregger, M. Stampanoni, M. Unser, "Constrained Regularized Reconstruction of X-Ray-DPCI Tomograms with Weighted-Norm," *Optics Express*, vol. 21, no. 26, pp. 32340-32348, 2013.




# Link with "Total variation" of Rudin-Osher

"Total variation of function" $\neq$ "total variation of a measure"

- Total variation in 1D: $\mathrm{TV}(f) = \sup_{\|\varphi\|_\infty \le 1} \langle \mathrm{D}f, \varphi \rangle = \|\mathrm{D}f\|_{\mathcal{M}}$

➟ perfect equivalence
(with L=D)

- Usual total variation in 2D: $\mathrm{TV}(f) = \sup_{\|\boldsymbol{\varphi}\|_\infty \le 1} \langle \boldsymbol{\nabla} f, \boldsymbol{\varphi} \rangle$

Problem: $\boldsymbol{\nabla} = (\partial_x, \partial_y)$ is not a scalar operator

$L_1$ version of the 2D total variation:

$$\mathrm{TV}(f) = \int_{\mathbb{R}^2} |\boldsymbol{\nabla} f(x,y)| \mathrm{d}x\mathrm{d}y, = \frac{1}{4} \int_0^{2\pi} \|\mathrm{D}_{\boldsymbol{u}_\theta} f\|_{\mathcal{M}} \, \mathrm{d}\theta$$

➟ angular averaging
(rotation invariance)

$\mathrm{D}_{\boldsymbol{u}_\theta} f = \langle \boldsymbol{u}_\theta, \boldsymbol{\nabla} f \rangle$: directional derivative of $f$ along $\boldsymbol{u}_\theta = (\cos\theta, \sin\theta)$

Present theory explains the regularisation effect of $\|\mathrm{D}_{\boldsymbol{u}_\theta} f\|_{\mathcal{M}}$