Representer theorems for convex regularization

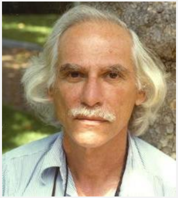
C. Boyer, Y. de Castro, A. Chambolle, V. Duval, A. Flinth, F. de Gournay, P. Weiss

Beijing, 21/04/2018

Introduction

The (older) heroes of this talk

Constantin Caratheodory (1873-1950)



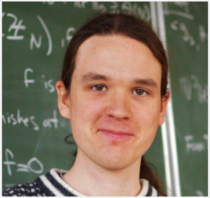
Lester Dubins (1920-2010)

Victor Klee (1925-2007)

And S.I. Zuhovickiĭ, G. Choquet, S. Fisher, J. Jerome...

Introduction

The (younger) heroes of this talk

Axel Flinth (1992-)

Vincent Duval (1983-)

Frédéric de Gournay (1980-)

And D. Donoho, V. Chandrasekaran etal, R. Chan,...

Introduction

Inverse problems

Let $u \in \mathcal{B}$, denote a signal from a vector space $\mathcal{B}$ (finite or infinite).
We are given a finite number m of corrupted linear measurements:

$$y = P(Au),$$

where

- $A : \mathcal{B} \to \mathbb{R}^m$ is defined by

$$(Au)_i = \langle a_i, u \rangle, a_i \in \mathcal{B}^*$$

- $P : \mathbb{R}^m \to \mathbb{R}^m$ is a perturbation operator (e.g. quantization, additive noise, modulus...).

Problem

How can we retrieve an approximation $\hat{u}$ of u knowing y and A?

Introduction

Example 1: Photography

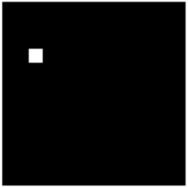
On a conventional camera:

$$a_i(\cdot) = h(\cdot - x_i)$$

where h is a function localized around 0 and x_i denotes a pixel center.

$$y_m = \Bigg\langle$$

,



$$\Bigg\rangle$$

Introduction

Example 2: Tomography

In tomography a_i allows measuring line integrals.

$y_m = \Big\langle$  ,  $\Big\rangle$

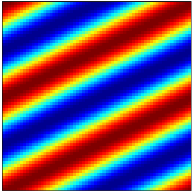
Introduction

Example 3: MRI

In MRI the functions a_i are complex exponentials.

$y_m = \Big\langle$

,



$\Big\rangle$

Introduction - Quadratic regularization

A critical issue

Regularization is critical whenever $\dim(\mathcal{B}) > m$.

Introduction - Quadratic regularization

A critical issue

Regularization is critical whenever $\dim(\mathcal{B}) > m$.

Tikhonov regularization (before 1943)

When $\mathcal{B}$ is a Hilbert space, we can solve:

$$\inf_{u \in \mathcal{B}} \frac{1}{2}\|Au - y\|_2^2 + \|Lu\|_{L^2}^2,$$

where $L : \mathcal{B} \to L^2$ is a linear operator (e.g. the derivative)

- ✔ Solutions given by linear systems.
- ✔ Sometimes solution of a finite dimensional problem yields an infinite dimensional solution (RKHS).
- × Typically restricts $\mathcal{B}$ to Hilbert spaces such as $W^{n,2}$.

Introduction - Quadratic regularization

A critical issue

Regularization is critical whenever $\dim(\mathcal{B}) > m$.

Tikhonov regularization (before 1943)

When $\mathcal{B}$ is a Hilbert space, we can solve:

$$\inf_{u\in\mathcal{B}} \frac{1}{2}\|Au - y\|_2^2 + \|Lu\|_{L^2}^2,$$

where $L : \mathcal{B} \to L^2$ is a linear operator (e.g. the derivative)

- ✔ Solutions given by linear systems.
- ✔ Sometimes solution of a finite dimensional problem yields an infinite dimensional solution (RKHS).
- × Typically restricts $\mathcal{B}$ to Hilbert spaces such as $W^{n,2}$.
- × Solutions live in a fixed subspace that depends on A and L only:

$$u^\star = \sum_{i=1}^{m} \alpha_i \psi_i + u_K, \quad \text{where } u_K \in \ker(L). \tag{1}$$

A first representer theorem.

Introduction - More recent approaches

Analysis formulation (before 1973)

$$\inf_{u\in\mathcal{B}} f_y(Au) + \|Lu\|_{\mathcal{M}},$$

- $L : \mathcal{B} \to \mathcal{M}$ is a linear operator (e.g. the derivative).
- $\mathcal{M}$ is the space of Radon measures.
- $f_y : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is a data fitting term.

Synthesis formulation (before 1973)

$$\inf_{\mu\in\mathcal{M}} f_y(AD\mu) + \|\mu\|_{\mathcal{M}},$$

where $D : \mathcal{M} \to \mathcal{B}$ is a linear operator called dictionary.

The estimate of $\hat{u}$ is given by $\hat{u} = D\hat{\mu}$.

S.D. Fisher and J.W. Jerome.
Spline solutions to l1 extremal problems in one and several variables.
Journal of Approximation Theory, 13(1):73–83, 1975.

Introduction - More recent approaches

Other popular examples

Nonnegative least squares:

$$\inf_{u\in B, u\geq 0} \frac{1}{2}\|Au-y\|_2^2.$$

Nuclear norm minimization:

$$\inf_{u\in\mathcal{B}, Au=y} \|u\|_*.$$

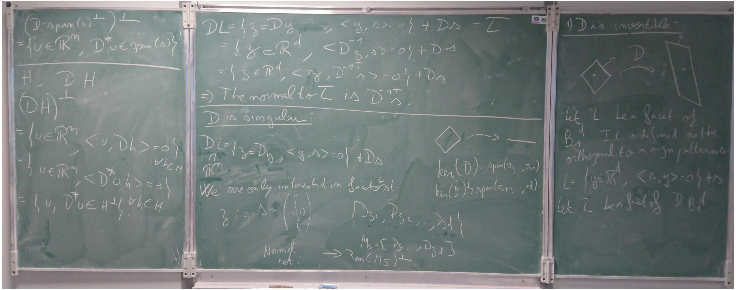
Plenty of such examples scattered in the literature.

The question tackled today

Can we derive representer theorems for problems of the form:

$$\inf_{u\in\mathcal{B}} f_y(Au)+R(u), \text{ where } R \text{ is convex?}$$

Part I: The main theoretical results



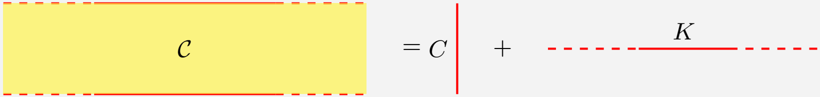
Preliminaries

Decomposition of a convex set

Let $\mathcal{C}$ denote a (linearly closed) convex set in $\mathcal{B}$.

We can decompose $\mathcal{C} = K + C$, where

- $K = \mathrm{Lin}(\mathcal{C})$ is the lineality space of $\mathcal{C}$,
- C is a linearly closed set that contains no line.



Convex gauge

The gauge of $\mathcal{C}$ is defined by:

$$R_{\mathcal{C}}(u) = \inf_{\lambda \geq 0, u \in \lambda \mathcal{C}} \lambda$$

Preliminaries

Our setting

I will focus on the properties of the minimizers of:

$$\inf_{u \in \mathcal{B}} f_y(Au) + R_{\mathcal{C}}(u).$$

where $f_y : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is either:

- a convex closed function.
- an arbitrary (nonconvex function).

We assume that the set of minimizers $\hat{U}$ is non empty.

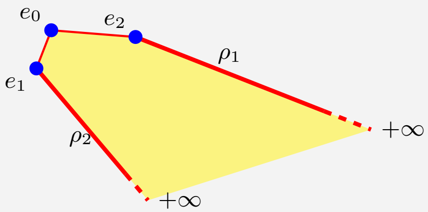
Preliminaries

Carathéodory - Klee (1957)

Let C denote a linearly closed convex set that contains no line in dimension m.

Then any point $u \in C$ can be expressed either as:

- A convex combination of $m+1$ points in $\mathrm{Ext}(C)$.
- A convex combination of m points in $\mathrm{Ext}(C) \cup \mathrm{Ray}(C)$.



V. Klee.
Extremal structure of convex sets.
Archiv der Mathematik, 8(3):234–240, August 1957.

Preliminaries

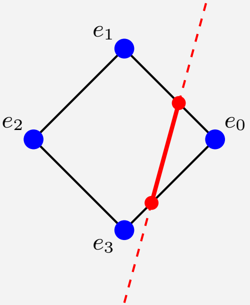
Dubins - Klee (1963)

Let C denote a linearly closed convex set that contains no line.
Let H denote an affine space of co-dimension m.

Then the extreme points and extreme rays of $C \cap H$ can be expressed as:

- A convex combination of $m + 1 + j$ points in $\text{Ext}(C)$.
- A convex combination of $m + j$ points in $\text{Ext}(C) \cup \text{Ray}(C)$.

Where $j = 0$ for the extreme points and $j = 1$ for the extreme rays.



Main results

A representer theorem: the nonconvex case (New result)

Consider the problem:

$$t^\star = \inf_{u \in \mathcal{B}} f_y(Au) + R_{\mathcal{C}}(u),$$

where f_y is an arbitrary function. Assume that at least one solution exists.

Then there exists a solution $\hat{u}$ of the form:

$$\hat{u} = \sum_{i=1}^{m+z} \alpha_i \psi_i + u_K,$$

where

- $u_K \in \mathrm{Lin}(\mathcal{C})$.
- $\psi_i \in \mathrm{Ext}(C) \cup \mathrm{Ray}(C)$ are the atoms of C.
- $z \leq 1_{t^\star = 0} - \dim(AK)$.

The bound is tight.

Main results

A representer theorem: the convex case (New result)

Consider the problem:

$$t^\star = \inf_{u \in \mathcal{B}} f_y(Au) + R_{\mathcal{C}}(u),$$

where f_y is either strictly convex or the indicator of a convex, linearly closed set. Assume that at least one solution exists.

Then the extreme points and rays of the solution set $\hat{U}$ are of the form:

$$\hat{u} = \sum_{i=1}^{m+z} \alpha_i \psi_i + u_K,$$

where

- $u_K \in \mathrm{Lin}(\mathcal{C})$.
- $\psi_i \in \mathrm{Ext}(C) \cup \mathrm{Ray}(C)$ are the atoms of C.
- $z \leq 1_{t^\star=0} + j + -\dim(AK)$, with $j = 0$ for extreme points and $j = 1$ for extreme rays.

Main results

The (rough) proof

Let $u^\star$ denote a solution and $t^\star = R_{\mathcal{C}}(u^\star)$. Consider the problem:

$$\inf_{u \in \mathcal{B}, Au = Au^\star} R_{\mathcal{C}}(u)$$

Any solution $\hat{u}$ is a solution of the original problem and satisfies $R_{\mathcal{C}}(\hat{u}) = t^\star$.
So $\hat{U} = H \cap D$, where:

$$H = \{u \in \mathcal{B}, Au = Au^\star\}$$

and

$$D = \{u \in C, R_{\mathcal{C}}(u) \leq t^\star\}.$$

Applying Klee's theorem (on $H \cap D$ quotiented by K), we get a complete description of this subset.
We can gain 1 point since the solutions live on the boundary of $\mathcal{C}$.

Part II: Examples of applications

Applications

ℓ_1 and total variation minimization

Consider the problems:

$$\inf_{u \in \mathbb{R}^n} f_y(Au) + \|u\|_1$$

or

$$\inf_{u \in \mathcal{M}} f_y(Au) + \|u\|_{\mathcal{M}}$$

There is at least one solution are m sparse:

$$\hat{u} = \sum_{i=1}^{m} \alpha_i \delta_{z_i}.$$

S.C. Chen, D. Donoho, and M. Saunders.
Atomic decomposition by basis pursuit.
SIAM review, 43(1):129–159, 2001.

D.L. Donoho.
Compressed sensing.
IEEE T. Inf. Theory, 52(4):1289–1306, 2006.

E. Candès and C. Fernandez-Granda.
Towards a mathematical theory of super-resolution.
Communications on Pure and Applied Mathematics, 67(6):906–956, 2014.

Applications

Nonnegative constraints

Consider the problem:

$$\min_{u \in \mathbb{R}^n_+} \frac{1}{2} \|Au - y\|_2^2$$

Then the extreme points and rays of the solution set are m sparse.

Don't use ℓ^1 when looking for sparse nonnegative signals!

D. Donoho and J. Tanner.
Sparse nonnegative solution of underdetermined linear equations by linear programming.
P. Nat. Acad. Sci. USA, 102(27):9446–9451, 2005.

A. Eftekhari, J. Tanner, A. Thompson, B. Toader, and H. Tyagi.
Sparse non-negative super-resolution-simplified and stabilised.
arXiv preprint arXiv:1804.01490, 2018.

Applications

Analysis priors - finite dimension

Let $L \in \mathbb{R}^{m \times n}$ denote a linear mapping.
Consider the problem:

$$\min_{u \in \mathbb{R}^n} f_y(Au) + \|Lu\|_1.$$

Then:

- If L is surjective, at least one solution can be written as:

$$\hat{u} = \sum_{i=1}^{m} \alpha_i L^+ \delta_{z_i} + u_K, u_K \in \ker(L).$$

- If L is not surjective, then there is a combinatorial explosion of the extreme points:

$$\#\mathrm{Ext}(\{u \in \mathbb{R}^n, \|Lu\|_1 \leq 1\}) \leq 2^{m-n+1} C_m^{m-n+1}$$

  Finding the vertices is the convex hull problem.

Applications

Analysis priors - infinite dimension (Old and new results)

Let $L : \mathcal{B} \to \mathcal{M}$ denote a linear and surjective mapping (plus some technical assumptions), where

$$\mathcal{B} = \{u \in \mathcal{D}', Lu \in \mathcal{M}, \|u\|_K < \infty\}.$$

Consider the problem:

$$\inf_{u \in \mathcal{B}} f_y(Au) + \|Lu\|_{\mathcal{M}}.$$

Then at least one solution is of the form:

$$\hat{u} = \sum_{i=1}^{m} \alpha_i L^+(\delta_{z_i}) + u_K.$$

S.D. Fisher and J.W. Jerome.
Spline solutions to l1 extremal problems in one and several variables.
Journal of Approximation Theory, 13(1):73–83, 1975.

M. Unser, J. Fageot, and John P. Ward.
Splines are universal solutions of linear inverse problems with generalized tv regularization.
SIAM Review, 59(4):769–793, 2017.

A. Flinth and P. Weiss.
Exact solutions of infinite dimensional total-variation regularized problems.
arXiv preprint arXiv:1708.02157, 2017.

Applications

Analysis priors - Biharmonic approximation (New result)

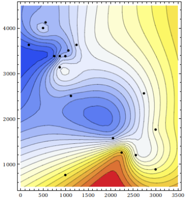
Solve

$$\inf_{u\in\mathcal{B}} \frac{1}{2}\sum_{i=1}^{m}(u(x_i)-y_i)^2 + \|\Delta\Delta u\|_{\mathcal{M}}.$$

Letting $\psi(x) = \|x\|^2\log(\|x\|)$, we get a solution of the form:

$$\hat{u} = \sum_{i=1}^{m}\alpha_i\psi(\cdot - z_i) + u_K,$$

is a polyharmonic spline, with u_K a polynomial of degree 1.



Polyharmonic splines are used for data interpolation

Applications

Analysis priors - Biharmonic approximation (New result)

Solve

$$\inf_{u\in\mathcal{B}} \frac{1}{2}\sum_{i=1}^{m}(u(x_i)-y_i)^2 + \|\Delta\Delta u\|_{\mathcal{M}}.$$

Letting $\psi(x) = \|x\|^2\log(\|x\|)$, we get a solution of the form:

$$\hat{u} = \sum_{i=1}^{m}\alpha_i\psi(\cdot - z_i) + u_K,$$

is a polyharmonic spline, with u_K a polynomial of degree 1.

The traditional approach

Usually, polyharmonic splines are appearing in the frame of RKHS.

$$\inf_{u\in H^2(\mathbb{R}^2)} \frac{1}{2}\sum_{i=1}^{m}(u(x_i)-y_i)^2 + \|\Delta u\|^2_{L^2(\mathbb{R}^d)}.$$

Applications

Total gradient variation (New result)

Consider the following problem:

$$\inf_{u \in BV(\mathbb{R}^d)} f_y(Ax) + \|Du\|_{\mathcal{M}},$$

then there exists a solution of the form:

$$\hat{u} = \sum_{i=1}^{m} \alpha_i \psi_i + c,$$

where c is a constant and

$$\psi_i = \mathbb{1}_{\omega_i}, \text{ where } \omega_i \text{ is a simple set.}$$

W.H. Fleming.
Functions with generalized gradient and generalized surfaces.
Annali di Matematica Pura ed Applicata, 44(1):93–103, 1957.

L. Ambrosio, V. Caselles, S. Masnou, and J.M. Morel.
Connected components of sets of finite perimeter and applications to image processing.
Journal of the European Mathematical Society, 3(1):39–92, 2001.

Applications

Other applications...

- Nuclear norm minimization ⇒ low rank.
- Linear, semi-definite and conic programming ⇒ sparse, low rank.
- Optimal transport ⇒ permutation matrices.
- Rank sparsity ball ⇒ low rank and sparse.
- ...

Some notes on computing

Representer theorems allow solving infinite dimensional problem exactly!

E. Candès and C. Fernandez-Granda.
Towards a mathematical theory of super-resolution.
Communications on Pure and Applied Mathematics, 67(6):906–956, 2014.

V. Duval and G. Peyré.
Exact support recovery for sparse spikes deconvolution.
Foundations of Computational Mathematics, 15(5):1315–1355, 2015.

A. Flinth and P. Weiss.
Exact solutions of infinite dimensional total-variation regularized problems.
arXiv preprint arXiv:1708.02157, 2017.

Some references

Thank you very much!

S.I. Zuhovickiĭ.
Remarks on problems in approximation theory.
Mat. Zbirnik KDU, pages 169–183, 1948.
(Ukrainian).

V. Klee.
Extremal structure of convex sets.
Archiv der Mathematik, 8(3):234–240, August 1957.

S.D. Fisher and J.W. Jerome.
Spline solutions to l1 extremal problems in one and several variables.
Journal of Approximation Theory, 13(1):73–83, 1975.

V. Chandrasekaran, B. Recht, P.A. Parrilo, and A.S. Willsky.
The convex geometry of linear inverse problems.
Found. Comp. Math., 12(6):805–849, 2012.

A. Flinth and P. Weiss.
Exact solutions of infinite dimensional total-variation regularized problems.
arXiv preprint arXiv:1708.02157, 2017.

C. Boyer, Y. de Castro, A. Chambolle, F. de Gournay, and P. Weiss.
Representer theorems for convex regularized inverse problems.
arXiv, 2018.