An atomic norm perspective on total variation regularization in image processing

Vincent Duval

INRIA Paris (MOKAPLAN) & U. Paris-Dauphine (CEREMADE)

Mathematics of Imaging Workshop # 1
February 6th, 2019

Joint work with

Claire Boyer

Antonin Chambolle

Yohann De Castro

Frédéric de Gournay

Pierre Weiss

Joint work with

 Claire Boyer

Antonin Chambolle 

 Yohann De Castro

Frédéric de Gournay 

 Pierre Weiss

Simultaneous with

 Kristian Bredies

Marcello Carioni 

Outline

1. Representer theorems in signal processing

2. The Dubins-Klee theorem

3. Application to Total Variation in Imaging

Summary

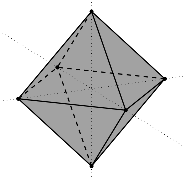
1. Representer theorems in signal processing

2. The Dubins-Klee theorem

3. Application to Total Variation in Imaging

A standard representer theorem

Let $\Phi : \mathbb{R}^P \to \mathbb{R}^M$, linear (surjective), and $y \in \mathbb{R}^M$.



The ℓ^1-minimization problem [Chen & Donoho'94]

$$\min_{x \in \mathbb{R}^P} \|x\|_1 \quad \text{s.t.} \quad \Phi x = y, \qquad (\mathcal{P})$$

admits a solution which is M-**sparse**.

☞ For M measurements some solution has at most M-**nonzero entries**.

Remarks:

- (Somewhat) Interesting if $M \ll P$
- Also holds for problems of the form

$$\min_{x \in \mathbb{R}^P} \lambda \|x\|_1 + g(\Phi x, y), \qquad (\mathcal{P}_\lambda)$$

  with $g(\cdot, y)$ convex.
- Other variants in linear programming, semi-definite programs, RKHS...

The Basis Pursuit for measures

Let $X \subset \mathbb{R}^d$, compact, and $\Phi\mu \stackrel{\text{def.}}{=} \left(\int_X \varphi_i(x)\mathrm{d}\mu(x)\right)_{1\leqslant i\leqslant M}$ for $\mu \in \mathcal{M}(X)$ and $\{\varphi_i\}_{i=1}^M \subset \mathscr{C}(X)$.
Consider

$$\min_{\mu\in\mathcal{M}(X)} \|\mu\|_{\mathcal{M}(X)} \quad \text{s.t.} \quad \Phi\mu = y, \qquad (\mathcal{P})$$

The **total variation of a measure** μ is

$$\|\mu\|_{\mathcal{M}(X)} = \sup\left\{\int_X \psi \mathrm{d}\mu; \psi \in \mathscr{C}(X), \|\psi\|_\infty \leqslant 1\right\}.$$

- If $\mu = \sum_{i=0}^N a_i\delta_{x_i}$, then $\|\mu\|_{\mathcal{M}(X)} = \sum_{i=0}^N |a_i|$.
- If $\mu = fd\mathcal{L}$, then $\|\mu\|_{\mathcal{M}(X)} = \int_X |f(t)|dt$.

Considered in [de Castro & Gamboa (12), Bredies & Pikkarainen (13), Candes & Fernandez-Granda (13), Recht et al. (12)] for deconvolution, frequency estimation, super-resolution,. . .

The Basis Pursuit for measures

Let $X \subset \mathbb{R}^d$, compact, and $\Phi\mu \overset{\text{def.}}{=} \left(\int_X \varphi_i(x)\mathrm{d}\mu(x)\right)_{1\leqslant i\leqslant M}$ for $\mu \in \mathcal{M}(X)$ and $\{\varphi_i\}_{i=1}^M \subset \mathscr{C}(X)$.
Consider

$$\min_{\mu\in\mathcal{M}(X)} \|\mu\|_{\mathcal{M}(X)} \quad \text{s.t.} \quad \Phi\mu = y, \qquad (\mathcal{P})$$

The **total variation of a measure** μ is

$$\|\mu\|_{\mathcal{M}(X)} = \sup\left\{\int_X \psi \mathrm{d}\mu; \psi \in \mathscr{C}(X), \|\psi\|_\infty \leqslant 1\right\}.$$

- If $\mu = \sum_{i=0}^N a_i \delta_{x_i}$, then $\|\mu\|_{\mathcal{M}(X)} = \sum_{i=0}^N |a_i|$.
- If $\mu = f d\mathcal{L}$, then $\|\mu\|_{\mathcal{M}(X)} = \int_X |f(t)| dt$.

…but it dates at least from the works of Krein and Beurling in the 1930's!

Representer theorems

$$\min_{\mu \in \mathcal{M}(X)} \|\mu\|_{\mathcal{M}(X)} \quad \text{s.t.} \quad \Phi\mu = y, \qquad (\mathcal{P})$$

- Zuhovickii (1948):
  There exists a solution to $\mathcal{P}$ of the form $\mu = \sum_{i=1}^{M} a_i \delta_{x_i}$.
- Fisher-Jerome (1973):
  The **extreme points** of $(\operatorname{argmin} \mathcal{P})$ are of the form $\mu = \sum_{i=1}^{M} a_i \delta_{x_i}$.

Representer theorems

$$\min_{\mu \in \mathcal{M}(X)} \|\mu\|_{\mathcal{M}(X)} \quad \text{s.t.} \quad \Phi\mu = y, \qquad (\mathcal{P})$$

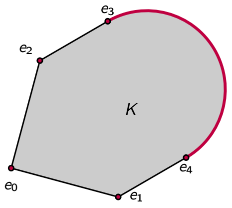
- Zuhovickii (1948):
  There exists a solution to $\mathcal{P}$ of the form $\mu = \sum_{i=1}^{M} a_i \delta_{x_i}$.
- Fisher-Jerome (1973):
  The **extreme points** of $(\operatorname{argmin} \mathcal{P})$ are of the form $\mu = \sum_{i=1}^{M} a_i \delta_{x_i}$.

x is an **extreme point** of the convex set K if for all x_1, $x_2 \in C$,

$$(x \in [x_1, x_2]) \Longrightarrow (x_1 = x_2 = x).$$

In a locally convex space Hausdorff space, if K is **compact**, the Krein-Milman theorem states that

$$K = \overline{\mathrm{conv}\,(\mathrm{extr}(K))}.$$



☞ Knowing the extreme points of $(\operatorname{argmin} \mathcal{P})$ gives access to **the full set of solutions**!

Representer theorems

$$\min_{\mu \in \mathcal{M}(X)} \|\mu\|_{\mathcal{M}(X)} \quad \text{s.t.} \quad \Phi\mu = y, \tag{$\mathcal{P}$}$$

- Zuhovickii (1948):
  There exists a solution to $\mathcal{P}$ of the form $\mu = \sum_{i=1}^{M} a_i \delta_{x_i}$.
- Fisher-Jerome (1973):
  The **extreme points** of $(\operatorname{argmin} \mathcal{P})$ are of the form $\mu = \sum_{i=1}^{M} a_i \delta_{x_i}$.
- Fisher-Jerome (1973), Unser et al. (2017), Flinth-Weiss (2017):
  If $L : \mathcal{F} \to \mathcal{M}(X)$ is onto, the **extreme points** of the solution set to

$$\min_{u \in \mathcal{F}} \|Lu\|_{\mathcal{M}(X)} \text{ such that } \Phi u = y \tag{$\mathcal{P}^{\mathcal{F}}$}$$

  are of the form $u = \sum_{i=1}^{M} a_i L^{+} \delta_{x_i} + v$, where $v \in \ker L$.

Consequences

Theory: The total variation promotes Dirac masses.

Numerics: Solve an infinite dimensional problem on your computer!

- Look for a solution of the form $\sum_{i=1}^{M} a_i \delta_{x_i}$,
- Parametrize the problem with $(a_i, x_i)_{1 \leqslant i \leqslant M} \in \mathbb{R}^{2M}$,
- See the conditional gradient/Frank-Wolfe method in [Bredies & Pikkarainen (13), Boyd et al. (17), Denoyelle et al. (18)].

How general is this phenomenon?
Can we do the same for the total variation of the gradient?

Consequences

Theory: The total variation promotes Dirac masses.

Numerics: Solve an infinite dimensional problem on your computer!

- Look for a solution of the form $\sum_{i=1}^{M} a_i \delta_{x_i}$,
- Parametrize the problem with $(a_i, x_i)_{1 \leqslant i \leqslant M} \in \mathbb{R}^{2M}$,
- See the conditional gradient/Frank-Wolfe method in [Bredies & Pikkarainen (13), Boyd et al. (17), Denoyelle et al. (18)].

How general is this phenomenon?
Can we do the same for the total variation of the gradient?

Key observation

- The 1-sparse vectors $\{\pm(0, \ldots, 0, 1, 0, \ldots, 0)^T\}$ are the extreme points of

$$\left\{x \in \mathbb{R}^P \; ; \; \|x\|_1 \leqslant 1\right\}.$$

- The Dirac masses $\{\pm\delta_x\}_{x \in X}$ are the extreme points of

$$\left\{\mu \in \mathcal{M}(X) \; ; \; \|\mu\|_{\mathcal{M}(X)} \leqslant 1\right\}.$$

A first try

$$\min_{u \in E} R(u) \quad \text{s.t.} \qquad \Phi u = y \qquad (\mathcal{P})$$

Theorem

Let E be a Banach space, $R : E \to \mathbb{R}_+$ be a convex, positively homogeneous function, $\Phi : E \to \mathbb{R}^M$ linear

Then every extreme point of $(\operatorname{argmin} \mathcal{P})$ is a convex combination of at most M extreme points of $\{R \leqslant \min \mathcal{P}\}$.

A first try

$$\min_{u \in E} R(u) \quad \text{s.t.} \qquad \Phi u = y \tag{$\mathcal{P}$}$$

Theorem

Let E be a Banach space, $R : E \to \mathbb{R}_+$ be a convex, positively homogeneous function such that $\{R \leqslant 1\}$ is compact, $\Phi : E \to \mathbb{R}^M$ linear
Then every extreme point of $(\operatorname{argmin} \mathcal{P})$ is a convex combination of at most M extreme points of $\{R \leqslant \min \mathcal{P}\}$.

A first try

$$\min_{u \in E} R(u) \quad \text{s.t.} \qquad \Phi u = y \tag{$\mathcal{P}$}$$

Theorem

Let E be a ~~Banach~~ locally convex vector space, $R : E \to \mathbb{R}_+$ be a convex, positively homogeneous function such that $\{R \leqslant 1\}$ is compact, $\Phi : E \to \mathbb{R}^M$ linear continuous.
Then every extreme point of $(\operatorname{argmin} \mathcal{P})$ is a convex combination of at most M extreme points of $\{R \leqslant \min \mathcal{P}\}$.

The choice of the topology seems crucial!

Summary

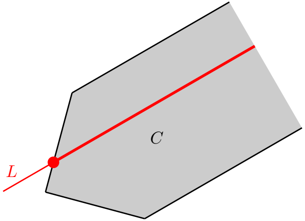
1. Representer theorems in signal processing

2. The Dubins-Klee theorem

3. Application to Total Variation in Imaging

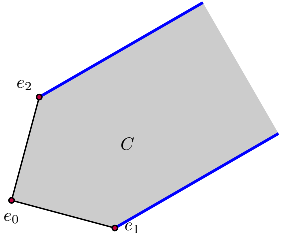
Let E be a vector space, $C \subseteq E$ a convex set.

C is **linearly closed** if for any line L, the endpoints of $L \cap C$ belong to C



A half-line $\rho \subseteq C$ is an **extreme ray of C if**

$$\forall x \in \rho, \forall x_1, x_2 \in C, (x \in]x_1, x_2[) \Rightarrow]x_1, x_2[\subset \rho.$$

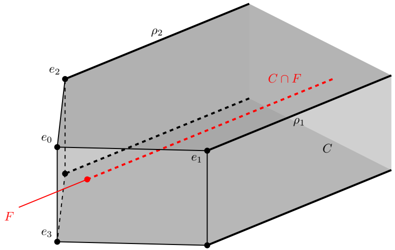


The Dubins-Klee theorem

Theorem (Dubins (1962), Klee (1963))

*Let E be a **vector space**, $C \subseteq E$ be convex, **linearly closed** and which contains no line.*
If $F \subseteq E$ is an affine space of codimension M, then each extreme point of $C \cap F$ is

- *a convex combination of (at most) $M + 1$ extreme points of C,*
- *or a convex combination of (at most) M points, each an extreme point or in an extreme ray of C.*

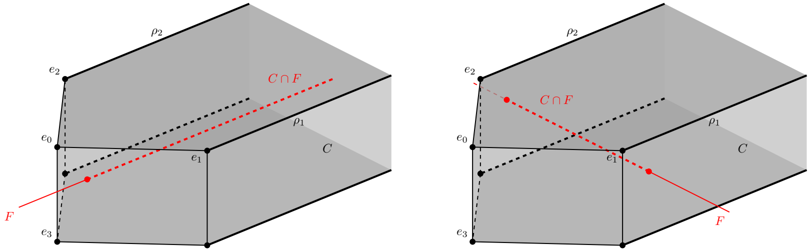


The Dubins-Klee theorem

Theorem (Dubins (1962), Klee (1963))

*Let E be a **vector space**, $C \subseteq E$ be convex, **linearly closed** and which contains no line.*
If $F \subseteq E$ is an affine space of codimension M, then each extreme point of $C \cap F$ is

- *a convex combination of (at most) $M + 1$ extreme points of C,*
- *or a convex combination of (at most) M points, each an extreme point or in an extreme ray of C.*



The Dubins-Klee theorem

Theorem (Dubins (1962), Klee (1963))

Let E be a ***vector space****, $C \subseteq E$ be convex,* ***linearly closed*** *and which contains no line.*
If $F \subseteq E$ is an affine space of codimension M, then each extreme point of $C \cap F$ is

- *a convex combination of (at most) $M + 1$ extreme points of C,*
- *or a convex combination of (at most) M points, each an extreme point or in an extreme ray of C.*

Remark:

- The non-emptyness of $\mathrm{extr}(C \cap F)$ is implicitly assumed,
- but, if $\mathrm{extr}(C \cap F) \neq \emptyset$, the theorem ensures that $\mathrm{extr}(C) \neq \emptyset$.
- Deals with the unbounded case!

The Dubins-Klee theorem

Theorem (Dubins (1962), Klee (1963))

Let E be a ***vector space****, $C \subseteq E$ be convex,* ***linearly closed*** *and which contains no line.*
If $F \subseteq E$ is an affine space of codimension M, then each extreme point of $C \cap F$ is

- *a convex combination of (at most) $M + 1$ extreme points of C,*
- *or a convex combination of (at most) M points, each an extreme point or in an extreme ray of C.*

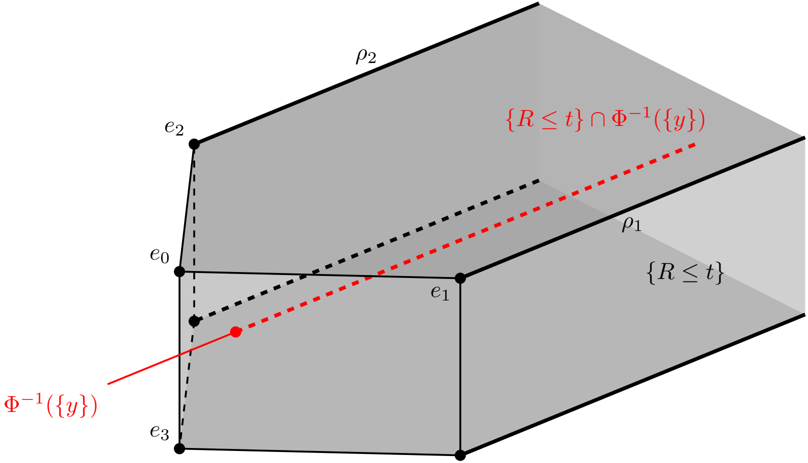
Example:
Let $E = \mathcal{M}(X)$, $F = \{\mu \in \mathcal{M}(X)\ ;\ \Phi\mu = y\}$.

$$\text{Choose } C = \left\{\mu \in \mathcal{M}(X)\ ;\ \|\mu\|_{\mathcal{M}(X)} \leqslant \min(\mathcal{P})\right\},$$

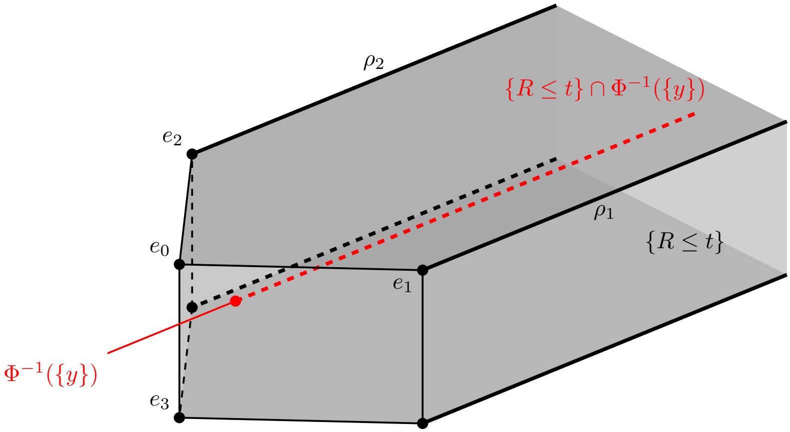
$$\text{hence} \quad \operatorname{argmin}(\mathcal{P}) = C \cap F.$$

Any of its extreme points can be written as $\mu = \sum_{i=1}^{M+1} a_i \delta_{x_i}$.

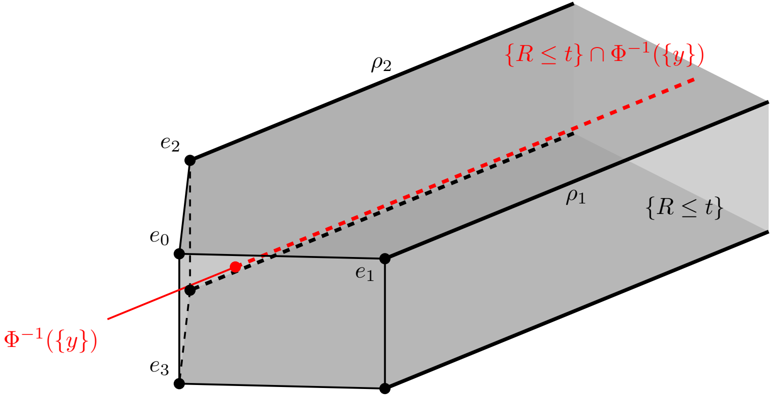
The case of convex optimization



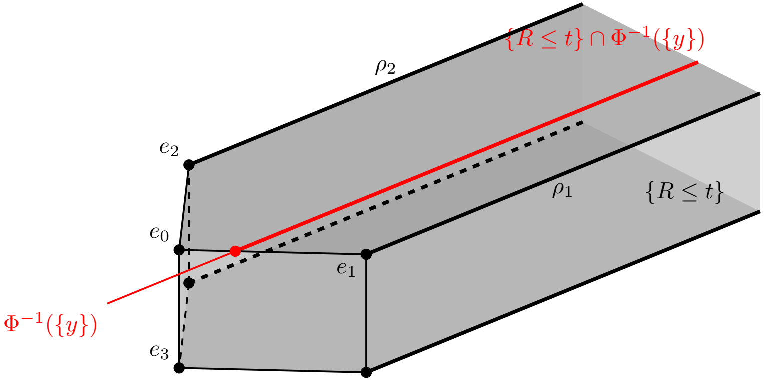
The case of convex optimization



The case of convex optimization

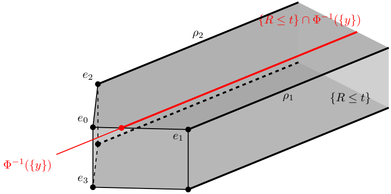


The case of convex optimization



The main theorem

$$\min_{u \in E} R(u) \quad \text{s.t.} \quad \Phi u = y \qquad (\mathcal{P})$$



$M = 2$

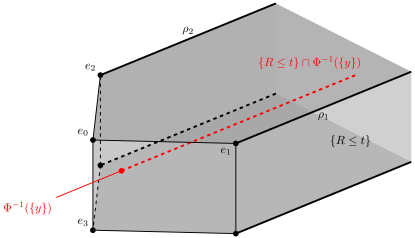
Theorem (BCCDGW ('18))

Let E be a vector space, $R : E \to \mathbb{R} \cup \{+\infty\}$ a convex function, and $\Phi : E \to \mathbb{R}^M$ linear such that $\operatorname{argmin}(\mathcal{P})$ *is nonempty. Assume that* $\{R \leqslant \min(\mathcal{P})\}$ *is linearly closed and contains no line.*
If $\min(\mathcal{P}) > \inf_E R$, *then each extreme point of* $\operatorname{argmin}(\mathcal{P})$ *is*

- *a convex combination of (at most) M extreme points of* $\{R \leqslant \min(\mathcal{P})\}$,
- *or a convex combination of (at most) $M - 1$ points, each an extreme point or in an extreme ray of* $\{R \leqslant \min(\mathcal{P})\}$.

The main theorem

$$\min_{u \in E} R(u) \quad \text{s.t.} \qquad \Phi u = y \qquad (\mathcal{P})$$



$M = 2$

Theorem (BCCDGW ('18))

Let E be a vector space, $R : E \to \mathbb{R} \cup \{+\infty\}$ a convex function, and $\Phi : E \to \mathbb{R}^M$ linear such that $\operatorname{argmin}(\mathcal{P})$ *is nonempty. Assume that* $\{R \leqslant \min(\mathcal{P})\}$ *is linearly closed and contains no line.*
If $\min(\mathcal{P}) = \inf_E R$, *then each extreme point of* $\operatorname{argmin}(\mathcal{P})$ *is*

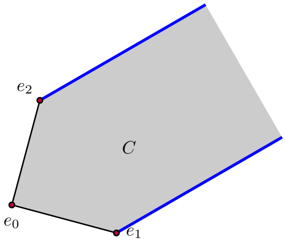
- *a convex combination of (at most) $M+1$ extreme points of* $\{R \leqslant \min(\mathcal{P})\}$,
- *or a convex combination of (at most) M points, each an extreme point or in an extreme ray of* $\{R \leqslant \min(\mathcal{P})\}$.

Idea of the proof

1. Show that each extreme point of $\operatorname{argmin} \mathcal{P}$ belongs to **a face** of $\{R \leqslant \min \mathcal{P}\}$ **of dimension (at most)** $M-1$.
2. Use Klee's extension of Carathéodory's theorem unbounded sets.

If $C \subset \mathbb{R}^d$ is closed, convex, and contains no line, then any point of C is a convex combination of

- (at most) $d+1$ extreme points of C,
- (at most) d points, each an extreme point or a point in an extreme ray of C.



Examples

$$\min_{\mu \in \mathcal{M}(X)} \|\mu\|_{\mathcal{M}(X)} \quad \text{s.t.} \quad \Phi\mu = y$$

The extreme points can be written as $\mu = \sum_{i=1}^{M} a_i \delta_{x_i}$.

Examples

$$\min_{\mu \in \mathcal{M}(X)} \|\mu\|_{\mathcal{M}(X)} \quad \text{s.t.} \quad \Phi\mu = y$$

The extreme points can be written as $\mu = \sum_{i=1}^{M} a_i \delta_{x_i}$.

$$\min_{\mu \in \mathcal{M}(X)} \iota_{\mathcal{M}_+(X)}(\mu) \quad \text{s.t.} \quad \Phi\mu = y$$

The extreme points can be written as $\mu = \sum_{i=1}^{M} a_i \delta_{x_i}$, $a_i \geqslant 0$.

Examples

$$\min_{\mu\in\mathcal{M}(X)} \|\mu\|_{\mathcal{M}(X)} \quad \text{s.t.} \quad \Phi\mu = y$$

The extreme points can be written as $\mu = \sum_{i=1}^{M} a_i \delta_{x_i}$.

$$\min_{\mu\in\mathcal{M}(X)} \iota_{\mathcal{M}_+(X)}(\mu) \quad \text{s.t.} \quad \Phi\mu = y$$

The extreme points can be written as $\mu = \sum_{i=1}^{M} a_i \delta_{x_i}$, $a_i \geqslant 0$.

$$\min_{S\in\mathcal{S}^n(\mathbb{R})} \iota_{\mathcal{S}^n_+(\mathbb{R})}(S) \quad \text{s.t.} \quad \Phi\mu = y$$

The extreme points can be written as $S = \sum_{i=1}^{M} a_i v_i \otimes v_i$, $a_i \geqslant 0$.

Examples

$$\min_{\mu \in \mathcal{M}(X)} \|\mu\|_{\mathcal{M}(X)} \quad \text{s.t.} \quad \Phi\mu = y$$

The extreme points can be written as $\mu = \sum_{i=1}^{M} a_i \delta_{x_i}$.

$$\min_{\mu \in \mathcal{M}(X)} \iota_{\mathcal{M}_+(X)}(\mu) \quad \text{s.t.} \quad \Phi\mu = y$$

The extreme points can be written as $\mu = \sum_{i=1}^{M} a_i \delta_{x_i}$, $a_i \geqslant 0$.

$$\min_{S \in \mathcal{S}^n(\mathbb{R})} \iota_{\mathcal{S}_+^n(\mathbb{R})}(S) \quad \text{s.t.} \quad \Phi\mu = y$$

The extreme points can be written as $S = \sum_{i=1}^{M} a_i v_i \otimes v_i$, $a_i \geqslant 0$.
Useless if $M \geqslant n$!!

Understanding the faces of C

If $S \in \mathcal{S}^n(\mathbb{R})$ has rank r, $S = \sum_{i=1}^{r} a_i v_i \otimes v_i$

Trick

$S \in \mathcal{S}^n_+(\mathbb{R})$ has rank (at most) r iff it belongs to a face of dimension $d = \frac{1}{2} r(r+1)$ of $\mathcal{S}^n_+(\mathbb{R})$.

- We obtain $S = \sum_{i=1}^{r} a_i v_i \otimes v_i$, with $r \leqslant \frac{1}{2}\left(\sqrt{8M+1} - 1\right) \leqslant M$.
- Coincides with known results [Barvinok'95].

Understanding the faces of C

If $S \in \mathcal{S}^n(\mathbb{R})$ has rank r, $S = \sum_{i=1}^{r} a_i v_i \otimes v_i$

Trick

$S \in \mathcal{S}_+^n(\mathbb{R})$ has rank (at most) r iff it belongs to a face of dimension $d = \frac{1}{2} r(r+1)$ of $\mathcal{S}_+^n(\mathbb{R})$.

- We obtain $S = \sum_{i=1}^{r} a_i v_i \otimes v_i$, with $r \leqslant \frac{1}{2}\left(\sqrt{8M+1} - 1\right) \leqslant M$.
- Coincides with known results [Barvinok'95].

"**the extreme points of** $\operatorname{argmin} \mathcal{P}$ **belong to a face of** C **of dimension (at most)** $M - 1$".

☞ Understanding the faces of C sometimes provides sharper results than the plain Carathéodory/Klee theorem.

Summary

1. Representer theorems in signal processing

2. The Dubins-Klee theorem

3. Application to Total Variation in Imaging

Total Variation Image Recovery

The **total** (gradient) **variation** of u is

$$\int_{\mathbb{R}^d} |Du| \stackrel{\text{def.}}{=} \sup\left\{ \int_{\mathbb{R}^d} u(x)\mathrm{div}z(x)\mathrm{d}x \ ; \ z \in \mathscr{C}_c^\infty(\mathbb{R}^d), \|z\|_\infty \leqslant 1 \right\}.$$

Following [Rudin et al.'92, Chambolle & Lions '97,...], given sensing functions $\{\varphi_i\}_{i=1}^M \subset L^d(\mathbb{R}^d)$ and an observation $y \in \mathbb{R}^M$, solve

$$\min_{u \in E} \int_{\mathbb{R}^d} |Du| \quad \text{s.t.} \quad \Phi\mu = y \qquad (\mathcal{P})$$

where $E \stackrel{\text{def.}}{=} L^{d/(d-1)}(\mathbb{R}^d)$, and $\Phi u \stackrel{\text{def.}}{=} \left(\int_{\mathbb{R}^d} u(x)\varphi_i(x)\mathrm{d}x \right)_{1 \leqslant i \leqslant M}$.

Ground truth $\longrightarrow$ Measurement

Total Variation Image Recovery

The **total** (gradient) **variation** of u is

$$\int_{\mathbb{R}^d} |Du| \stackrel{\text{def.}}{=} \sup \left\{ \int_{\mathbb{R}^d} u(x)\mathrm{div} z(x)\mathrm{d}x \ ; \ z \in \mathscr{C}_c^\infty(\mathbb{R}^d), \|z\|_\infty \leqslant 1 \right\}.$$

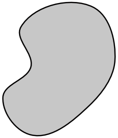
Following [Rudin et al.'92, Chambolle & Lions '97,...], given sensing functions $\{\varphi_i\}_{i=1}^M \subset L^d(\mathbb{R}^d)$ and an observation $y \in \mathbb{R}^M$, solve

$$\min_{u \in E} \int_{\mathbb{R}^d} |Du| \quad \text{s.t.} \quad \Phi\mu = y \tag{$\mathcal{P}$}$$

Theorem ([Fleming 1957, Ambrosio et al.'01])

The extreme points of $\left\{ u \in E \ ; \ \int_{\mathbb{R}^d} |Du| \leqslant 1 \right\}$ *are the functions of the form* $u = \pm \mathbb{1}_E / P(E)$, *where* E *is a* ***simple set****.*

Simple sets are the "simply connected" sets in the measure theoretic sense.

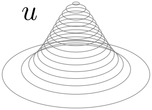


Note: see also [Bredies & Carioni'18] for the case of a bounded domain Ω.

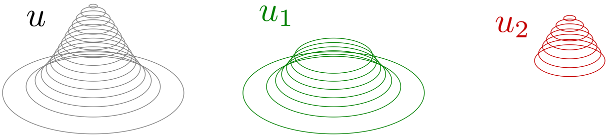
Idea of the proof

(see for instance [Bredies & Carioni'18])

Coarea formula: $\int_{\mathbb{R}^d} |Du| = \int_{-\infty}^{+\infty} P(\{u \geqslant t\})\mathrm{d}t$

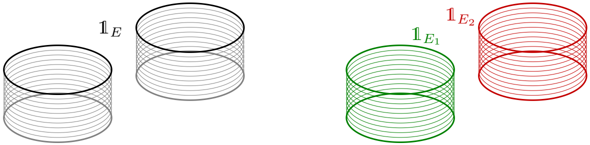


1. The function u can only have **one non-trivial level set**
   Otherwise, write $u = \theta \frac{u_1}{\int_{\mathbb{R}^d} |Du_1|} + (1-\theta) \frac{u_2}{\int_{\mathbb{R}^d} |Du_2|}$ with $\theta = \int_{\mathbb{R}^d} |Du_1| = 1 - \int_{\mathbb{R}^d} |Du_2|$.



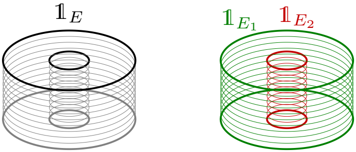
$\Rightarrow$ contradicts u being an extreme point. Hence $u = \pm \frac{\mathbb{1}_E}{P(E)}$.

2. The set E has to be "connected". Otherwise, write $\frac{\mathbb{1}_E}{P(E)} = \theta \frac{\mathbb{1}_{E_1}}{P(E_1)} + (1-\theta)\frac{\mathbb{1}_{E_2}}{P(E_2)}$



with $\theta = \frac{P(E_1)}{P(E)} = 1 - \frac{P(E_2)}{P(E)}$ since $P(E_1) + P(E_2) = P(E)$. Contradiction.

3. E cannot have "holes". Otherwise, write $\frac{\mathbb{1}_E}{P(E)} = \theta \frac{\mathbb{1}_{E_1}}{P(E_1)} + (1-\theta)\frac{(-\mathbb{1}_{E_2})}{P(E_2)}$



with $\theta = \frac{P(E_1)}{P(E)} = 1 - \frac{P(E_2)}{P(E)}$ since $P(E_1) + P(E_2) = P(E)$.
Contradiction. $\square$

Representation of the solutions

Consequence: The extreme points of argmin($\mathcal{P}$) are *sums of at most M indicators of simple sets*,

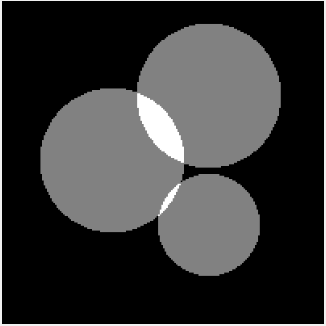
$$u = \sum_{i=1}^{M} a_i \mathbb{1}_{E_i}.$$

Yet another explanation of the **staircasing phenomenon!**
(see also [Nikolova'00, Jalalzai'16,...])

Representation of the solutions

Consequence: The extreme points of $\operatorname{argmin}(\mathcal{P})$ are *sums of at most M indicators of simple sets*,

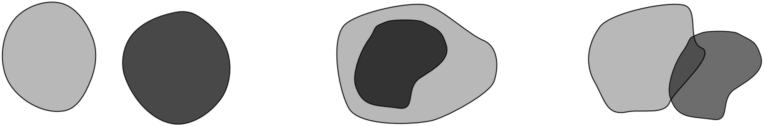
$$u = \sum_{i=1}^{M} a_i \mathbb{1}_{E_i}.$$



Measurement functions φ_i

Solution u

The faces of the TV unit ball

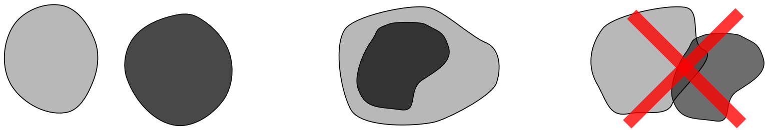


The faces of the TV unit ball

Proposition (BCDDGW'19)

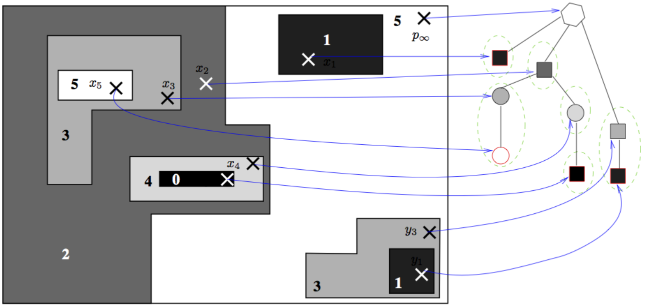
Let $A, B \subset \mathbb{R}^d$ be simple sets such that $[\frac{\mathbb{1}_A}{P(B)}, \frac{\mathbb{1}_B}{P(B)}]$ is a face of $\{\int |Du| \leqslant 1\}$. Then

- $|A \setminus B| = 0$ or $|B \setminus A| = 0$,
- or $|A \cap B| = 0$ and $\mathcal{H}^{N-1}(\partial^* A \cap \partial^* B) = 0$.



The tree of shapes of an image

Decomposition of an images into a tree of shapes [Monasse & Guichard'00, Ballester et al.'03].



[Ballester et al.'03]

Theorem (BCDDGW'19)

If u is an extreme point of $\operatorname{argmin}(\mathcal{P})$, *its tree of shapes has at most* $M+1$ *nodes.*

Conclusion

- A simple and general representation principle for inverse problems
- Understanding the extreme points (or the faces) of a regularizer yields powerful a description of the solution.
- Makes it possible to numerically solve infinite dimensional problems!
- Perspectives:
  - Understand other regularizers (eg. TGV)
  - Derive numerical methods for "off-the-grid" total variation minimization.

Thank you for your attention!

On Representer Theorems and Convex Regularization, C. Boyer, A. Chambolle, Y. De Castro, V. Duval, F. de Gournay, P. Weiss. 2018. *Accepted for publication at SIAM J. on Optimization*

Sparsity of solutions for variational inverse problems with finite-dimensional data, K. Bredies, M. Carioni. 2018. *arXiV preprint 1809.05045*

Thank you for your attention!

The unbounded case

Theorem (BCCDGW ('18))

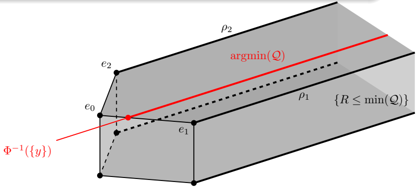
If argmin($\mathcal{Q}$) *is nonempty, linearly closed and contains no line, and*

$$\min(\mathcal{Q}) > \inf_E R,$$

then each of its extreme points is

- *a convex combination of (at most) M extreme points of $\{R \leqslant \min(\mathcal{Q})\}$,*
- *or a convex combination of (at most) $M-1$ points, each an extreme point or in an extreme ray of $\{R \leqslant \min(\mathcal{Q})\}$.*

The proof relies on a theorem of [Klee 1963] for the representation of convex sets.



$M = 2$