Regularization and Representer Theorems

on a topology-free diet

Yohann De Castro (CERMICS)

Labex Day, 4th December 2018

ON REPRESENTER THEOREMS AND CONVEX REGULARIZATION[*]

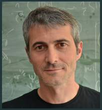
CLAIRE BOYER[*], ANTONIN CHAMBOLLE[†], YOHANN DE CASTRO[‡§], VINCENT DUVAL[§¶], FRÉDÉRIC DE GOURNAY[‖], AND PIERRE WEISS[#]

Abstract. We establish a general principle which states that regularizing an inverse problem with a convex function yields solutions which are convex combinations of a small number of *atoms*. These atoms are identified with the extreme points and elements of the extreme rays of the regularizer level sets. An extension to a broader class of quasi-convex regularizers is also discussed. As a side result, we characterize the minimizers of the total gradient variation, which was still an unresolved problem.

Key words. Inverse problems, Convex regularization, Representer theorem, Vector space, Total variation

AMS subject classifications. 52A05, 49N45, 46E27

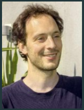
C. Boyer



A. Chambolle

V. Duval

F. de Gournay



P. Weiss

École des Ponts
ParisTech

1. Inverse Problems Regularization
2. Lineality Space, Extreme Rays
3. A Representer Theorem
4. Examples of Applications

Inverse Problems Regularization

Inverse problem: recover $u \in E$ from $y \in \mathbb{R}^m$ through a linear operator $\Phi : E \to \mathbb{R}^m$ perturbed by an operator $P : \mathbb{R}^m \to \mathbb{R}^m$,

$$y = P(\Phi u)$$

where E is a (locally convex Hausdorff) vector space and $m \in \mathbb{N}$.

Inverse problem: recover $u \in E$ from $y \in \mathbb{R}^m$ through a linear operator $\Phi : E \to \mathbb{R}^m$ perturbed by an operator $P : \mathbb{R}^m \to \mathbb{R}^m$,

$$y = P(\Phi u)$$

where E is a (locally convex Hausdorff) vector space and $m \in \mathbb{N}$.

Regularization: One may consider

$$\inf_{u \in E} f(\Phi u) + R(u), \tag{$\mathscr{P}$}$$

where $R : E \to \mathbb{R} \cup \{+\infty\}$ convex function called *regularizer* and f arbitrary function (convex or non-convex) called *data fitting term*.

One can be interested in [Scholkopf and Smola, 2001]

$$\min_{u\in\mathbb{R}^m} \frac{1}{2}\left\|\Phi u - y\right\|_2^2 + \frac{1}{2}\left\|Lu\right\|_2^2,$$

where $\Phi \in \mathbb{R}^{m\times n}$, $L \in \mathbb{R}^{p\times n}$ s.t. $\ker\Phi \cap \ker L = \{0\}$.

One can be interested in [Scholkopf and Smola, 2001]

$$\min_{u\in\mathbb{R}^m} \frac{1}{2}\|\Phi u - y\|_2^2 + \frac{1}{2}\|Lu\|_2^2,$$

where $\Phi \in \mathbb{R}^{m\times n}$, $L \in \mathbb{R}^{p\times n}$ s.t. $\ker\Phi \cap \ker L = \{0\}$.

Solutions are

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

with $u_K \in \ker(L)$ and $\psi_i = (\Phi^\top\Phi + L^\top L)^{-1}(\phi_i)$ denoting $\phi_i^\top \in \mathbb{R}^n$ the i-th row of Φ.

Lineality Space, Extreme Rays

Linearly Closed, Recession Cone and Lineality Space

Let E be a real vector space and let $C \subseteq E$ be a convex set.

Linearly Closed (resp. linearly bounded) as "Topology-free" Diet

Any intersection of C and a line of E is closed (resp. bounded) for the natural topology of the line.

Linearly Closed, Recession Cone and Lineality Space

Let E be a real vector space and let $C \subseteq E$ be a convex set.

Linearly Closed (resp. linearly bounded) as "Topology-free" Diet

Any intersection of C and a line of E is closed (resp. bounded) for the natural topology of the line.

Recession Cone, $\mathrm{rec}(C)$

Set of all $v \in E$ s.t. $C + \mathbb{R}_+^* v \subseteq C$. It is a convex cone.

École des Ponts ParisTech

Linearly Closed, Recession Cone and Lineality Space

Let E be a real vector space and let $C \subseteq E$ be a convex set.

Linearly Closed (resp. linearly bounded) as "Topology-free" Diet

Any intersection of C and a line of E is closed (resp. bounded) for the natural topology of the line.

Recession Cone, $\mathrm{rec}(C)$

Set of all $v \in E$ s.t. $C + \mathbb{R}_+^* v \subseteq C$. It is a convex cone.

Lineality Space, $\mathrm{lin}(C)$

$$\mathrm{lin}(C) := \mathrm{rec}(C) \cap (-\mathrm{rec}(C))$$

École des Ponts ParisTech

Extreme points and Extreme Rays

Extreme Points and Rays

- Extreme Points: points $p \in C$ s.t. $C \setminus \{p\}$ is convex;

École des Ponts ParisTech

Extreme points and Extreme Rays

Extreme Points and Rays

- Extreme Points: points $p \in C$ s.t. $C \setminus \{p\}$ is convex;
- Extreme Rays: rays $\rho \in C$ s.t. if $x, y \in C$ and $]x, y[$ intersects ρ, then $]x, y[\subset \rho$;

Extreme points and Extreme Rays

Extreme Points and Rays

- Extreme Points: points $p \in C$ s.t. $C \setminus \{p\}$ is convex;
- Extreme Rays: rays $\rho \in C$ s.t. if $x, y \in C$ and $]x, y[$ intersects ρ, then $]x, y[\subset \rho$;
- Faces $\mathscr{F}_C(p)$: Union of $\{p\}$ and all the open segments in C which have p as an inner point;

École des Ponts ParisTech

Extreme points and Extreme Rays

Extreme Points and Rays

- Extreme Points: points $p \in C$ s.t. $C \setminus \{p\}$ is convex;
- Extreme Rays: rays $\rho \in C$ s.t. if $x, y \in C$ and $]x, y[$ intersects ρ, then $]x, y[\subset \rho$;
- Faces $\mathscr{F}_C(p)$: Union of $\{p\}$ and all the open segments in C which have p as an inner point;

Faces description and Quotienting by lines

Denote W a supplement of $\mathrm{lin}(C)$ and $\widetilde{C} := C \cap W$ then

$$C = \widetilde{C} + \mathrm{lin}(C), \text{ and } \{\mathscr{F}_{\widetilde{C}}(p) + \mathrm{lin}(C)\}_{p \in \widetilde{C}} = \{\mathscr{F}_C(p)\}_{p \in C}$$

is the partition of C in elementary faces.

A Representer Theorem

Representer Theorem

Denote $t^\star$ the *optimal value* of (1) given by

$$\min_{u \in E} R(u) \quad \text{s.t.} \quad \Phi u = y\,, \tag{1}$$

$\mathscr{S}^\star$ its *solution set*, and $C^\star \overset{\text{def.}}{=} \left\{ u \in E \;:\; R(u) \leq t^\star \right\}$.

Representer Theorem

Denote $t^\star$ the *optimal value* of (1) given by

$$\min_{u \in E} R(u) \quad \text{s.t.} \quad \Phi u = y\,, \tag{1}$$

$\mathscr{S}^\star$ its *solution set*, and $C^\star \stackrel{\text{def.}}{=} \left\{ u \in E \ : \ R(u) \leq t^\star \right\}$.

Theorem ([Boyer et al., 2018])

If $\inf_E R < t^\star < +\infty$, $\mathscr{S}^\star$ *nonempty,* $C^\star$ *is linearly closed and contains no line, and* $p \in \mathscr{S}^\star$ *s.t.* *j is the dimension of the face $\mathscr{F}_{\mathscr{S}^\star}(p)$.*
Then *p belongs to a face of $C^\star$ with dimension at most $m+j-1$ and it can be written as a convex combination of*

- *$m+j$ extreme points of $C^\star$,*
- *or $m+j-1$ points of $C^\star$, each an extreme point of $C^\star$ or in an extreme ray of $C^\star$.*

École des Ponts ParisTech

On a figure

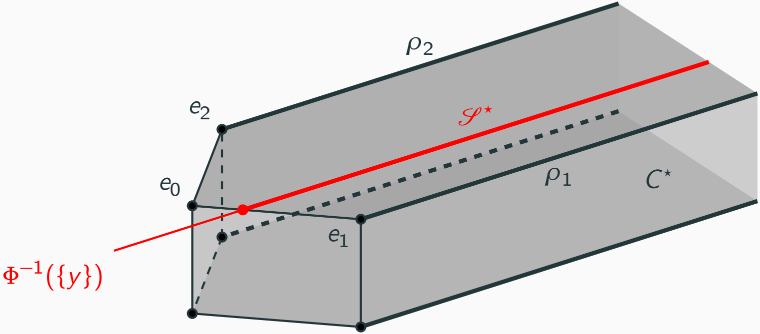


Figure 1: For $m = 2$ with $\mathscr{S}^\star = C^\star \cap \Phi^{-1}(\{y\})$ made of an extreme point and an extreme ray. The extreme point is a convex combination of $\{e_0, e_1\}$. Depending on their position, the points in the ray are a convex combination of $\{e_0, e_1, e_2\}$ or a pair of points, one in ρ_1 and the other in ρ_2.

École des Ponts ParisTech

Quotienting by lines on a figure

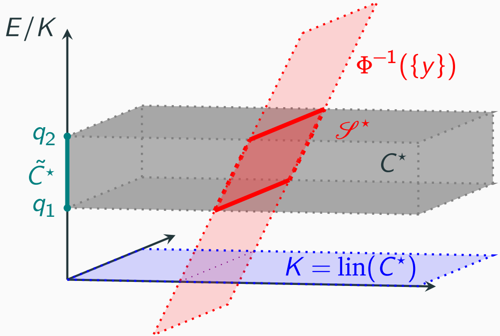


Figure 2: Quotienting by $K = \mathrm{lin}(C^\star)$ yields a level set $\widetilde{C^\star}$ with no line.

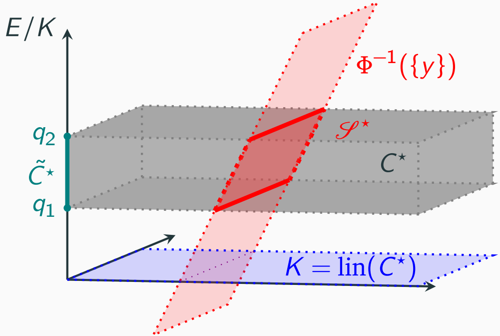


Figure 2: Quotienting by $K = \mathrm{lin}(C^\star)$ yields a level set $\widetilde{C^\star}$ with no line.

With $\tilde{q}_1, \dots, \tilde{q}_r \in \tilde{C}^\star$, $d \overset{\text{def.}}{=} \dim \Phi(K)$, $r \leq m + j - d \; (-1)$,

$$p = \sum_{i=1}^{r} \theta_i \underbrace{\psi_K^{-1}(\tilde{q}_i, 0)}_{q_i \in E} + u_K, \quad \text{where} \quad \theta_i \geq 0, \; \sum_{i=1}^{r} \theta_i = 1, \quad \text{and} \quad u_K \in K.$$

Examples of Applications

Linear Programming and the Moment Problem

$$\inf_{\substack{\mu \in \mathscr{M}_+(\Omega) \\ \Phi\mu = y}} \langle \psi, u \rangle. \tag{2}$$

with Ω compact metric space, $\mathscr{M}_+(\Omega)$ nonnegative Radon measures, ψ and $(\phi_i)_{1 \leq i \leq m}$ continuous.

École des Ponts ParisTech

$$\inf_{\substack{\mu \in \mathscr{M}_+(\Omega) \\ \Phi\mu = y}} \langle \psi, u \rangle. \tag{2}$$

with Ω compact metric space, $\mathscr{M}_+(\Omega)$ nonnegative Radon measures, ψ and $(\phi_i)_{1 \le i \le m}$ continuous.

Assume that the solution set (2) is nonempty. Then, its extreme points are m-sparse, i.e. of the form:

$$u = \sum_{i=1}^{m} \alpha_i \delta_{x_i}, \quad x_i \in \Omega, \alpha_i \geq 0.$$

École des Ponts ParisTech

The Total Variation ball

$$B_{\mathscr{M}} = \left\{ u \in \mathscr{M}(\Omega) \ : \ \|u\|_{\mathscr{M}} \leq 1 \right\}$$

with Ω open subset of $\mathbb{R}^d$ and $\mathscr{M}(\Omega)$ Radon measures. One has

$$\mathrm{ext}(B_{\mathscr{M}}) = \{\pm\delta_x, \ x \in \Omega\}$$

École des Ponts ParisTech

The Total Variation ball

$$B_{\mathcal{M}} = \left\{ u \in \mathcal{M}(\Omega) \ : \ \|u\|_{\mathcal{M}} \leq 1 \right\}$$

with Ω open subset of $\mathbb{R}^d$ and $\mathcal{M}(\Omega)$ Radon measures. One has

$$\mathrm{ext}(B_{\mathcal{M}}) = \{\pm\delta_x \, , \ x \in \Omega\}$$

Total variation regularized problems of the form:

$$\inf_{u \in \mathcal{M}} f(\Phi u) + \|u\|_{\mathcal{M}},$$

yield m-sparse solutions (under an existence assumption).

The Total Gradient Variation 1/2

For any locally integrable function u define

$$TV(u) \overset{\text{def.}}{=} \sup\left(\int u\mathrm{div}(\phi)\, dx, \phi \in C_c^1(\mathbb{R}^d)^d, \sup_{x\in\mathbb{R}^d} \|\phi(x)\|_2 \leq 1\right).$$

If finite then gradient Du is a Radon measure and

$$TV(u) = \int_{\mathbb{R}^d} |Du| = \|Du\|_{(\mathscr{M}(\mathbb{R}^d))^d}.$$

École des Ponts
ParisTech

The Total Gradient Variation 1/2

For any locally integrable function u define

$$TV(u) \stackrel{\text{def.}}{=} \sup\left(\int u\mathrm{div}(\phi)\,dx, \phi \in C_c^1(\mathbb{R}^d)^d, \sup_{x\in\mathbb{R}^d} \|\phi(x)\|_2 \leq 1\right).$$

If finite then gradient Du is a Radon measure and

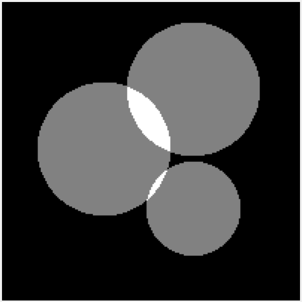
$$TV(u) = \int_{\mathbb{R}^d} |Du| = \|Du\|_{(\mathscr{M}(\mathbb{R}^d))^d}.$$

Theorem ([Fleming, 1957, Ambrosio et al., 2001])

Extreme points of the TV unit ball are indicators of simple sets normalized by their perimeter, i.e. $u = \pm\frac{1_F}{TV(1_F)}$, *where* F *is an indecomposable and saturated subset of* $\mathbb{R}^d$.

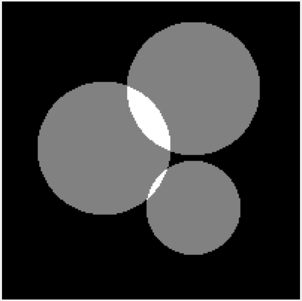
École des Ponts ParisTech

Minimizing the total variation s.t. a finite number of linear constraints can be expressed as a sum of a small number of indicators of simple sets. Explaining the stair-casing effect [Nikolova, 2000].



The Total Gradient Variation 2/2

Minimizing the total variation s.t. a finite number of linear constraints can be expressed as a sum of a small number of indicators of simple sets. Explaining the stair-casing effect [Nikolova, 2000].



References i

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

Boyer, C., Chambolle, A., De Castro, Y., Duval, V., De Gournay, F., and Weiss, P. (2018).
On representer theorems and convex regularization.
arXiv preprint arXiv:1806.09810.

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

Nikolova, M. (2000).
Local strong homogeneity of a regularized estimator.
SIAM Journal on Applied Mathematics, 61(2):633–658.

Scholkopf, B. and Smola, A. J. (2001).
***Learning with kernels: support vector machines, regularization, optimization, and beyond*.**
MIT press.

Questions?