Submodular Functions: from Discrete to Continuous Domains

Francis Bach

INRIA - Ecole Normale Supérieure

NIPS Workshop 2016

Submodular functions
From discrete to continuous domains
Summary

- **Which functions can be minimized in polynomial time?**
  - Beyond convex functions

Submodular functions
From discrete to continuous domains
Summary

- **Which functions can be minimized in polynomial time?**
  - Beyond convex functions
- **Submodular functions**
  - Not convex, ... but "equivalent" to convex functions
  - Usually defined on $\{0, 1\}^n$
  - Extension to continuous domains
  - *Application*: proximal operator for non-convex regularizers
- **Preprint available on ArXiv, second version (Bach, 2015)**

Submodularity for combinatorial optimization (see, e.g., Fujishige, 2005; Bach, 2013)

- **Definition**: $\forall x, y \in \{0,1\}^n$,

$$H(x) + H(y) \geqslant H(\max\{x,y\}) + H(\min\{x,y\})$$

  - NB: identification of $x \in \{0,1\}^n$ to $\{i,\ x_i = 1\} \subseteq \{1,\dots,n\}$

- **Examples**: cut functions, entropies, set covers, etc.

Submodularity for combinatorial optimization (see, e.g., Fujishige, 2005; Bach, 2013)

- **Definition**: $\forall x, y \in \{0,1\}^n$,

$$H(x) + H(y) \geqslant H(\max\{x,y\}) + H(\min\{x,y\})$$

  – NB: identification of $x \in \{0,1\}^n$ to $\{i,\ x_i = 1\} \subseteq \{1,\dots,n\}$

- **Examples**: cut functions, entropies, set covers, etc.

- **Minimization in polynomial time**

  – Reformulation as a convex problem through continuous extension

From discrete to continuous domains

- **Main insight**: $\{0, 1\}$ **is totally ordered!**

From discrete to continuous domains

- **Main insight**: $\{0,1\}$ **is totally ordered!**

- **Extension to** $\{0,\ldots,k-1\}$: $H : \{0,\ldots,k-1\}^n \to \mathbb{R}$

$$\forall x, y, \quad H(x) + H(y) \geqslant H(\min\{x,y\}) + H(\max\{x,y\})$$

  – Equivalent definition: with $(e_i)_{i\in\{1,\ldots,n\}}$ canonical basis of $\mathbb{R}^n$

$$\forall x, i \neq j, \quad H(x+e_i) + H(x+e_j) \geqslant H(x) + H(x+e_i+e_j)$$

  – See Lorentz (1953); Topkis (1978)

From discrete to continuous domains

- **Main insight**: $\{0,1\}$ **is totally ordered!**

- **Extension to** $\{0,\dots,k-1\}$: $H:\{0,\dots,k-1\}^n \to \mathbb{R}$

$$\forall x,y, \quad H(x)+H(y) \geqslant H(\min\{x,y\}) + H(\max\{x,y\})$$

  – Equivalent definition: with $(e_i)_{i\in\{1,\dots,n\}}$ canonical basis of $\mathbb{R}^n$

$$\forall x, i\neq j, \quad H(x+e_i)+H(x+e_j) \geqslant H(x) + H(x+e_i+e_j)$$

  – See Lorentz (1953); Topkis (1978)

- **Taylor expansion**:

  – $H(x+e_i)+H(x+e_j) \approx 2H(x) + \frac{\partial H}{\partial x_i} + \frac{\partial H}{\partial x_j} + \frac{1}{2}\frac{\partial^2 H}{\partial x_i^2} + \frac{1}{2}\frac{\partial^2 H}{\partial x_j^2}$

  – $H(x)+H(x+e_i+e_j) = 2H(x) + \frac{\partial H}{\partial x_i} + \frac{\partial H}{\partial x_j} + \frac{1}{2}\frac{\partial^2 H}{\partial x_i^2} + \frac{1}{2}\frac{\partial^2 H}{\partial x_j^2} + \frac{\partial^2 H}{\partial x_i \partial x_j}$

From discrete to continuous domains

- **Main insight**: $\{0,1\}$ **is totally ordered!**

- **Extension to** $\{0,\dots,k-1\}$: $H:\{0,\dots,k-1\}^n \to \mathbb{R}$

$$\forall x,y,\quad H(x)+H(y) \geqslant H(\min\{x,y\})+H(\max\{x,y\})$$

  – Equivalent definition: with $(e_i)_{i\in\{1,\dots,n\}}$ canonical basis of $\mathbb{R}^n$

$$\forall x, i\neq j,\quad H(x+e_i)+H(x+e_j) \geqslant H(x)+H(x+e_i+e_j)$$

  – See Lorentz (1953); Topkis (1978)

- **Taylor expansion**:

  – $H(x+e_i)+H(x+e_j) \approx 2H(x)+\frac{\partial H}{\partial x_i}+\frac{\partial H}{\partial x_j}+\frac{1}{2}\frac{\partial^2 H}{\partial x_i^2}+\frac{1}{2}\frac{\partial^2 H}{\partial x_j^2}$

  – $H(x)+H(x+e_i+e_j) = 2H(x)+\frac{\partial H}{\partial x_i}+\frac{\partial H}{\partial x_j}+\frac{1}{2}\frac{\partial^2 H}{\partial x_i^2}+\frac{1}{2}\frac{\partial^2 H}{\partial x_j^2}+\frac{\partial^2 H}{\partial x_i \partial x_j}$

From discrete to continuous domains

- **Main insight**: $\{0,1\}$ **is totally ordered!**
- **Extension to** $\{0,\dots,k-1\}$: $H:\{0,\dots,k-1\}^n \to \mathbb{R}$

$$\forall x,y, \quad H(x)+H(y) \geqslant H(\min\{x,y\}) + H(\max\{x,y\})$$

  – Equivalent definition: with $(e_i)_{i\in\{1,\dots,n\}}$ canonical basis of $\mathbb{R}^n$

$$\forall x, i\neq j, \quad H(x+e_i)+H(x+e_j) \geqslant H(x) + H(x+e_i+e_j)$$

  – See Lorentz (1953); Topkis (1978)

- **Generalization to all totally ordered sets**: $\mathcal{X}_i \subset \mathbb{R}$

  intervals $+\ H$ twice differentiable: $\forall x \in \prod_{i=1}^n \mathcal{X}_i, \quad \dfrac{\partial^2 H}{\partial x_i \partial x_j}(x) \leqslant 0$

A "new" class of continuous functions

- Assume each $\mathcal{X}_i \subset \mathbb{R}$ is a compact interval, and (for simplicity) H twice differentiable:

$$\textbf{Submodularity}: \ \forall x \in \prod_{i=1}^{n} \mathcal{X}_i, \quad \frac{\partial^2 H}{\partial x_i \partial x_j}(x) \leqslant 0$$

A "new" class of continuous functions

- Assume each $\mathcal{X}_i \subset \mathbb{R}$ is a compact interval, and (for simplicity) H twice differentiable:

$$\textbf{Submodularity}: \ \forall x \in \prod_{i=1}^{n} \mathcal{X}_i, \quad \frac{\partial^2 H}{\partial x_i \partial x_j}(x) \leqslant 0$$

- **Invariance** by
  - individual increasing smooth change of variables $H(\varphi_1(x_1), \dots, \varphi_n(x_n))$
  - adding arbitrary (smooth) separable functions $\sum_{i=1}^{n} v_i(x_i)$

A "new" class of continuous functions

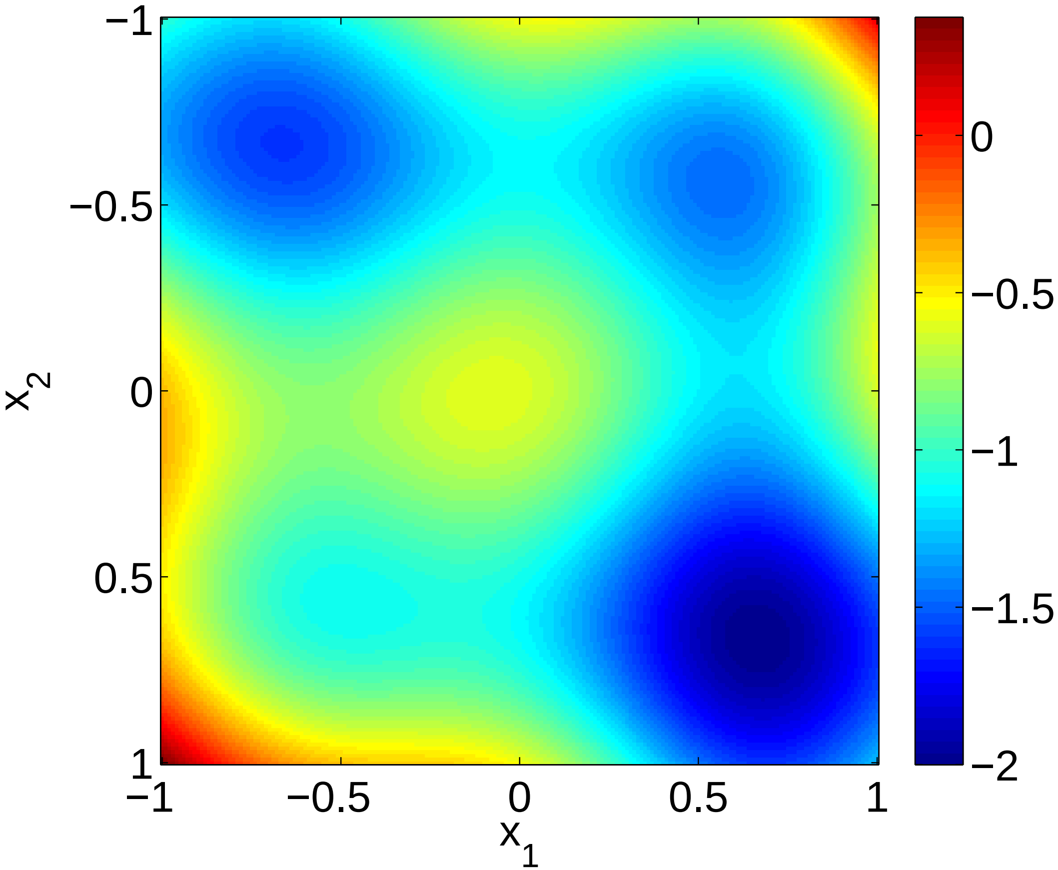
- Assume each $\mathcal{X}_i \subset \mathbb{R}$ is a compact interval, and (for simplicity) H twice differentiable:

$$\textbf{Submodularity}: \ \forall x \in \prod_{i=1}^n \mathcal{X}_i, \quad \frac{\partial^2 H}{\partial x_i \partial x_j}(x) \leqslant 0$$

- **Invariance** by
  - individual increasing smooth change of variables $H(\varphi_1(x_1), \ldots, \varphi_n(x_n))$
  - adding arbitrary (smooth) separable functions $\sum_{i=1}^n v_i(x_i)$

- **Examples**
  - Quadratic functions with Hessians with non-negative off-diagonal entries (Kim and Kojima, 2003)
  - $\psi(x_i - x_j)$, ψ convex; $\psi(x_1 + \cdots + x_n)$, ψ concave; $\log\det$, etc...
  - Monotone of order two (Carlier, 2003), Spence-Mirrlees condition (Milgrom and Shannon, 1994)

A “new” class of continuous functions



- Level sets of the submodular function $(x_1, x_2) \mapsto \frac{7}{20}(x_1 - x_2)^2 - e^{-4(x_1 - \frac{2}{3})^2} - \frac{3}{5}e^{-4(x_1 + \frac{2}{3})^2} - e^{-4(x_2 - \frac{2}{3})^2} - e^{-4(x_2 + \frac{2}{3})^2}$, with several local minima, local maxima and saddle points

Extensions to the space of product measures

View 1: thresholding cumulative distrib. functions

- Identify $x_i \in \mathcal{X}_i$ with the Dirac δ_{x_i} (a probability distribution on $\mathcal{X}_i$)

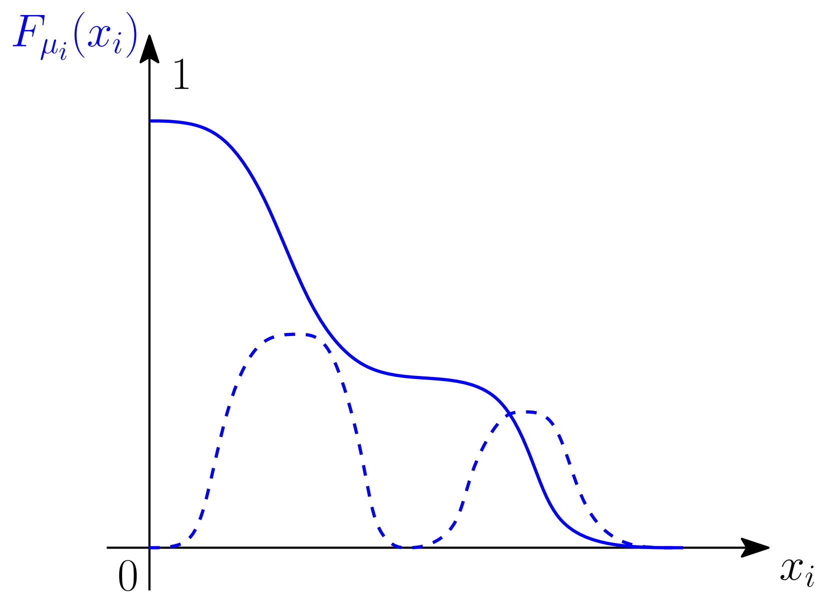
Extensions to the space of product measures

View 1: thresholding cumulative distrib. functions

- Given a probability distribution $\mu_i \in \mathcal{P}(\mathcal{X}_i)$
  - (reversed) cumulative distribution function $F_{\mu_i} : \mathcal{X}_i \to [0,1]$ as

$$F_{\mu_i}(x_i) = \mu_i\big(\{y_i \in \mathcal{X}_i, y_i \geqslant x_i\}\big) = \mu_i\big([x_i, +\infty)\big) \in [0,1]$$

  - and its "inverse": $F_{\mu_i}^{-1}(t) = \sup\{x_i \in \mathcal{X}_i,\ F_{\mu_i}(x_i) \geqslant t\} \in \mathcal{X}_i$



Extensions to the space of product measures
View 1: thresholding cumulative distrib. functions

- Given a probability distribution $\mu_i \in \mathcal{P}(\mathcal{X}_i)$
  - (reversed) cumulative distribution function $F_{\mu_i} : \mathcal{X}_i \to [0,1]$ as

$$F_{\mu_i}(x_i) = \mu_i\big(\{y_i \in \mathcal{X}_i, y_i \geqslant x_i\}\big) = \mu_i\big([x_i, +\infty)\big) \in [0,1]$$

  - and its "inverse": $F_{\mu_i}^{-1}(t) = \sup\{x_i \in \mathcal{X}_i,\ F_{\mu_i}(x_i) \geqslant t\} \in \mathcal{X}_i$

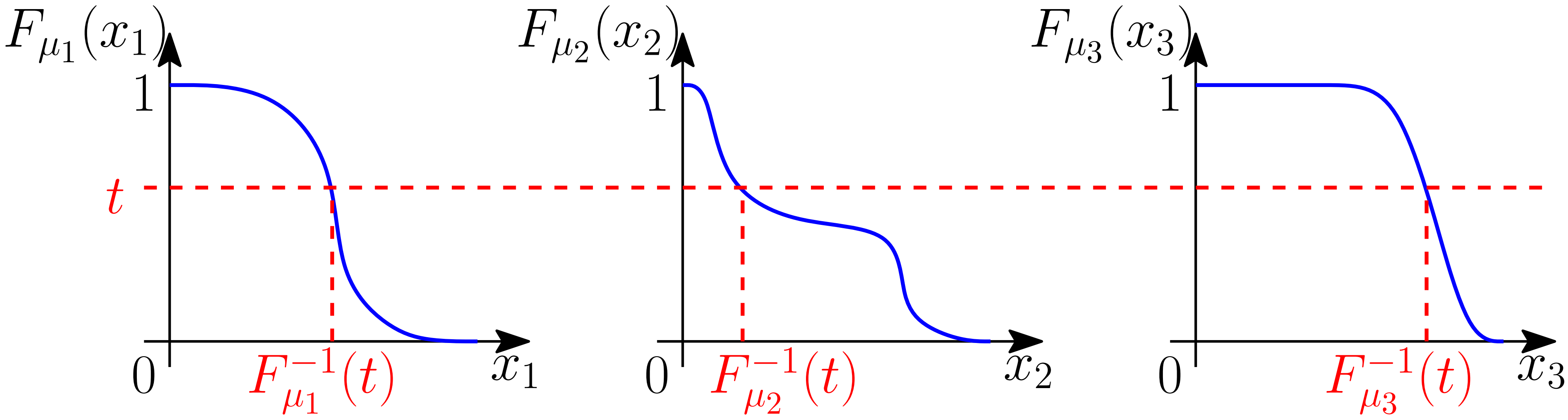
- **"Continuous" extension**

$$\forall \mu \in \prod_{i=1}^{n} \mathcal{P}(\mathcal{X}_i), \quad h(\mu_1, \dots, \mu_n) = \int_0^1 H\big[F_{\mu_1}^{-1}(t), \dots, F_{\mu_n}^{-1}(t)\big] dt$$

  - For finite sets, can be computed by sorting *all* values of $F_{\mu_i}(x_i)$
  - Equal to the "Lovász extension" when $\forall i$, $\mathcal{X}_i = \{0,1\}$

Extensions to the space of product measures

View 1: thresholding cumulative distrib. functions



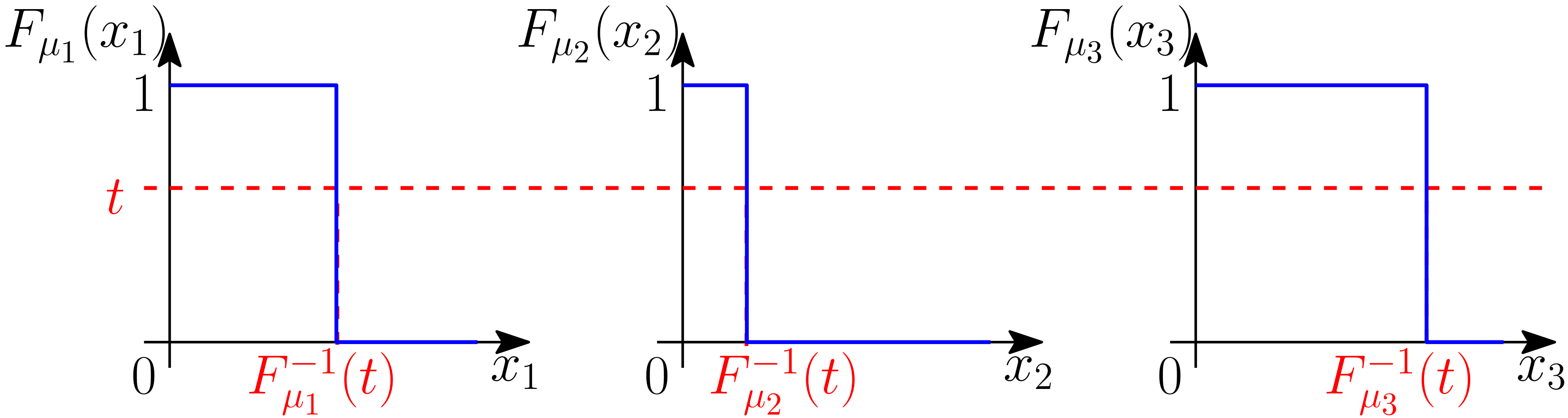
- **"Continuous" extension**

$$\forall \mu \in \prod_{i=1}^{n} \mathcal{P}(\mathcal{X}_i), \quad h(\mu_1, \dots, \mu_n) = \int_0^1 H\big[F_{\mu_1}^{-1}(t), \dots, F_{\mu_n}^{-1}(t)\big] dt$$

  – For finite sets, can be computed by sorting *all* values of $F_{\mu_i}(x_i)$
  – Equal to the "Lovász extension" when $\forall i$, $\mathcal{X}_i = \{0, 1\}$

Extensions to the space of product measures

View 1: thresholding cumulative distrib. functions



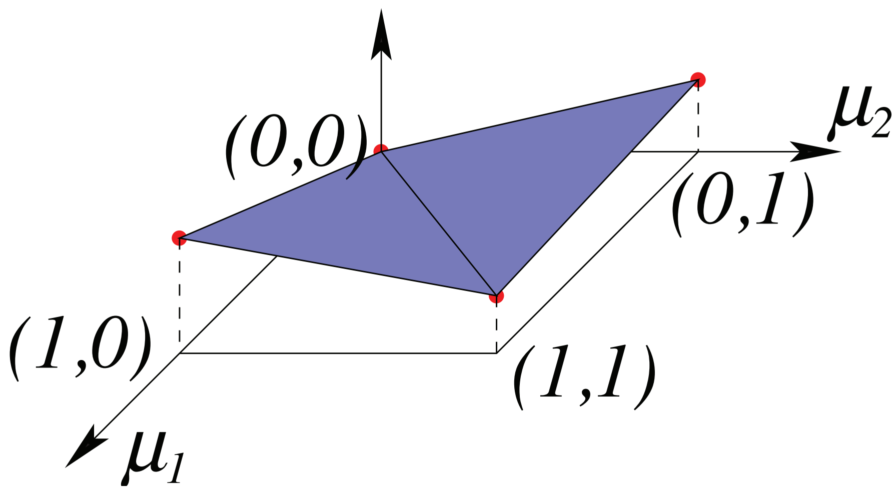
- **"Continuous" extension**

$$\forall \mu \in \prod_{i=1}^{n} \mathcal{P}(\mathcal{X}_i), \quad h(\mu_1, \dots, \mu_n) = \int_0^1 H\big[F_{\mu_1}^{-1}(t), \dots, F_{\mu_n}^{-1}(t)\big] dt$$

  – For finite sets, can be computed by sorting *all* values of $F_{\mu_i}(x_i)$
  – Equal to $H(x_1, \dots, x_n)$ when $\mu_i = \delta_{x_i}$ for all i

Extensions to the space of product measures
View 2: convex closure

- Given any function H on $\mathcal{X} = \prod_{i=1}^{n} \mathcal{X}_i$
  - Known value $H(x)$ for any "extreme points" of product measures (i.e., all Diracs δ_x at any $x \in \mathcal{X}$)
  - Convex closure $\tilde{h}$ = largest convex lower bound
  - Minimizing H and its convex closure $\tilde{h}$ is equivalent



Extensions to the space of product measures
View 2: convex closure

- Given any function H on $\mathcal{X} = \prod_{i=1}^n \mathcal{X}_i$
  - Known value $H(x)$ for any "extreme points" of product measures (i.e., all Diracs δ_x at any $x \in \mathcal{X}$)
  - Convex closure $\tilde{h}$ = largest convex lower bound
  - Minimizing H and its convex closure $\tilde{h}$ is equivalent

- Need to compute the Fenchel-Legendre bi-conjugate of

$$a : \mu \mapsto H(x) \text{ if } \mu = \delta_x \text{ for some } x \in \mathcal{X}, \text{ and } +\infty \text{ otherwise}$$

Extensions to the space of product measures
View 2: convex closure

- Given any function H on $\mathcal{X} = \prod_{i=1}^{n} \mathcal{X}_i$
  - Known value $H(x)$ for any "extreme points" of product measures (i.e., all Diracs δ_x at any $x \in \mathcal{X}$)
  - Convex closure $\tilde{h}$ = largest convex lower bound
  - Minimizing H and its convex closure $\tilde{h}$ is equivalent

- "Closed-form" formula: $\tilde{h}(\mu_1, \dots, \mu_n) = \inf_{\gamma \in \mathcal{P}(\mathcal{X})} \int_{\mathcal{X}} H(x) d\gamma(x)$,
  - Optimization with respect to all joint probability measures γ on $\mathcal{X}$ such that $\gamma_i(x_i) = \mu_i(x_i)$ (fixed marginals)

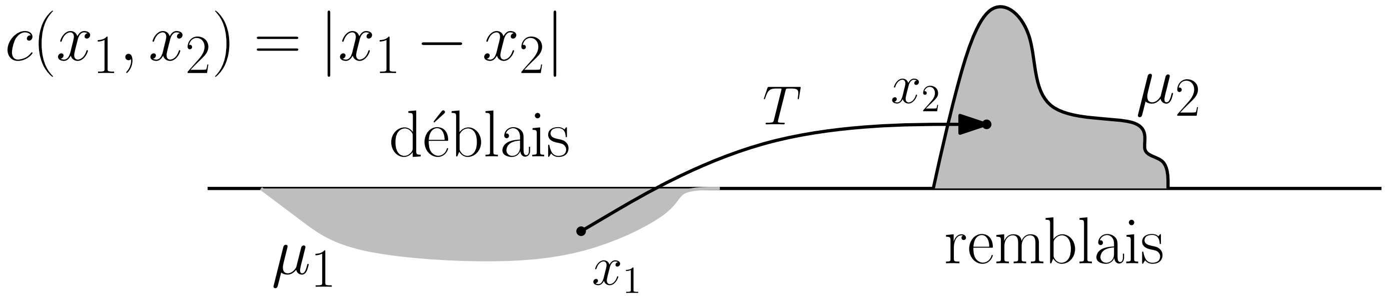
Extensions to the space of product measures
View 2: convex closure

- Given any function H on $\mathcal{X} = \prod_{i=1}^{n} \mathcal{X}_i$
  - Known value $H(x)$ for any "extreme points" of product measures (i.e., all Diracs δ_x at any $x \in \mathcal{X}$)
  - Convex closure $\tilde{h}$ = largest convex lower bound
  - Minimizing H and its convex closure $\tilde{h}$ is equivalent

- "Closed-form" formula: $\tilde{h}(\mu_1, \dots, \mu_n) = \inf_{\gamma \in \mathcal{P}(\mathcal{X})} \int_{\mathcal{X}} H(x) d\gamma(x),$
  - Optimization with respect to all joint probability measures γ on $\mathcal{X}$ such that $\gamma_i(x_i) = \mu_i(x_i)$ (fixed marginals)
  - Multi-marginal optimal transport

Optimal transport: from Monge to Kantorovich

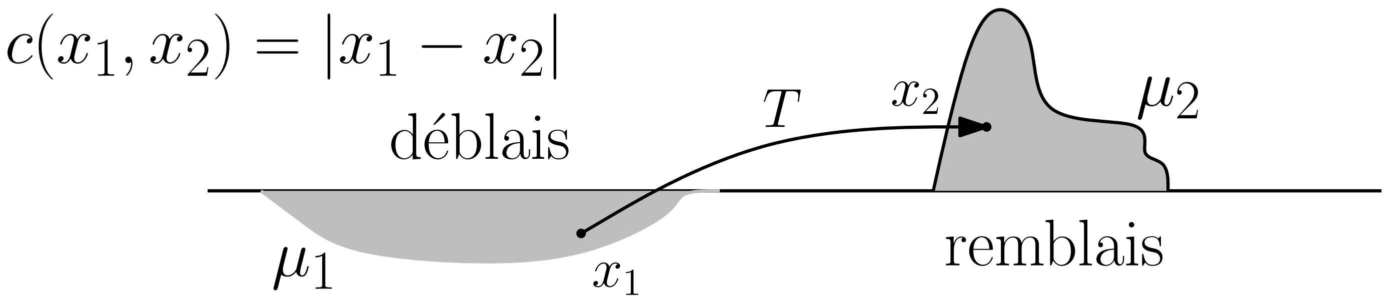
- **Monge formulation** ("La théorie des déblais et des remblais", 1781)
  - Transforming a measure μ_1 to μ_2 that (a) preserves local mass and (b) minimize transportation cost $\int_{\mathcal{X}_1} c(x_1, T(x_1)) d\mu_1(x_1)$



  - Optimal transport map T may not always exists
  - Discrete case: earth's mover distance

Optimal transport: from Monge to Kantorovich

- **Monge formulation** ("La théorie des déblais et des remblais", 1781)
  - Transforming a measure μ_1 to μ_2 that (a) preserves local mass and (b) minimize transportation cost $\int_{\mathcal{X}_1} c(x_1, T(x_1)) d\mu_1(x_1)$



  - Optimal transport map T may not always exists
  - Discrete case: earth's mover distance

- **Kantorovich formulation (1942)**
  - Convex relaxation on space of probability measures $\gamma \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$
  - Prescribed marginals $\gamma_1 = \mu_1$ and $\gamma_2 = \mu_2$
  - Minimum cost $\int_{\mathcal{X}_1 \times \mathcal{X}_2} c(x_1, x_2) d\gamma(x_1, x_2)$

Optimal transport: from two to multiple marginals

- **Kantorovich formulation (1942)**
  - Convex relaxation on space of probability measures $\gamma \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$
  - Prescribed marginals $\gamma_1 = \mu_1$ and $\gamma_2 = \mu_2$
  - Minimum cost $\int_{\mathcal{X}_1 \times \mathcal{X}_2} c(x_1, x_2) d\gamma(x_1, x_2)$

Optimal transport: from two to multiple marginals

- **Kantorovich formulation (1942)**
  - Convex relaxation on space of probability measures $\gamma \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$
  - Prescribed marginals $\gamma_1 = \mu_1$ and $\gamma_2 = \mu_2$
  - Minimum cost $\int_{\mathcal{X}_1 \times \mathcal{X}_2} c(x_1, x_2) d\gamma(x_1, x_2)$

- **Properties**
  - Monge formulation with distribution of $(x_1, T(x_1))$
  - Wasserstein distance between measures with $c(x_1, x_2) = |x_1 - x_2|^p$
  - See Villani (2008); Santambrogio (2015)

Optimal transport: from two to multiple marginals

- **Kantorovich formulation (1942)**
  - Convex relaxation on space of probability measures $\gamma \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$
  - Prescribed marginals $\gamma_1 = \mu_1$ and $\gamma_2 = \mu_2$
  - Minimum cost $\int_{\mathcal{X}_1 \times \mathcal{X}_2} c(x_1, x_2) d\gamma(x_1, x_2)$

- **Properties**
  - Monge formulation with distribution of $(x_1, T(x_1))$
  - Wasserstein distance between measures with $c(x_1, x_2) = |x_1 - x_2|^p$
  - See Villani (2008); Santambrogio (2015)

- **Extension to multiple marginals**
  - Minimize $\int_{\mathcal{X}} H(x) d\gamma(x_1, \dots, x_n)$ with respect to all prob. measures γ on $\mathcal{X}$ such that $\gamma_i(x_i) = \mu_i(x_i)$ for all $i \in \{1, \dots, n\}$

Extensions to the space of product measures
Combining the two views

- **View 1: thresholding cumulative distribution functions**

  + closed form computation for any H, always an extension

  − not convex

- **View 2: convex closure**

  + convex for any H, allows minimization of H

  − not computable, may not be an extension

Extensions to the space of product measures
Combining the two views

- **View 1: thresholding cumulative distribution functions**
  - \+ closed form computation for any H, always an extension
  - − not convex
- **View 2: convex closure**
  - \+ convex for any H, allows minimization of H
  - − not computable, may not be an extension
- **Submodularity**
  - The two views are equivalent
  - Direct proof through optimal transport
  - All results from submodular set-functions go through

Kantorovich optimal transport in one dimension

- **Theorem** (Carlier, 2003): If H is submodular, then

$$\inf_{\gamma \in \mathcal{P}(\mathcal{X})} \int_{\mathcal{X}} H(x) d\gamma(x) \text{ such that } \forall i, \gamma_i = \mu_i$$

is equal to $\displaystyle\int_0^1 H\big[F_{\mu_1}^{-1}(t), \dots, F_{\mu_n}^{-1}(t)\big] dt$

Kantorovich optimal transport in one dimension

- **Theorem** (Carlier, 2003): If H is submodular, then

$$\inf_{\gamma \in \mathcal{P}(\mathcal{X})} \int_{\mathcal{X}} H(x) d\gamma(x) \text{ such that } \forall i, \gamma_i = \mu_i$$

is equal to $\displaystyle\int_0^1 H\big[F_{\mu_1}^{-1}(t), \dots, F_{\mu_n}^{-1}(t)\big] dt$

"$\Leftrightarrow$" **Optimal transport with one-dimensional distributions and submodular cost is obtained in closed form**

– See Villani (2008); Santambrogio (2015)

Submodular functions
Links with convexity (Bach, 2015)

1. H **is submodular if and only if** h **is convex**

2. **If** H **is submodular, then**

$$\min_{x \in \prod_{i=1}^{n} \mathcal{X}_i} H(x) = \min_{\mu \in \prod_{i=1}^{n} \mathcal{P}(\mathcal{X}_i)} h(\mu)$$

3. **If** H **is submodular, then a subgradient of** h **at any** μ **may be computed by a "greedy algorithm"**

Submodular functions
Links with convexity (Bach, 2015)

1. H **is submodular if and only if** h **is convex**

2. **If** H **is submodular, then**

$$\min_{x \in \prod_{i=1}^{n} \mathcal{X}_i} H(x) = \min_{\mu \in \prod_{i=1}^{n} \mathcal{P}(\mathcal{X}_i)} h(\mu)$$

3. **If** H **is submodular, then a subgradient of** h **at any** μ **may be computed by a "greedy algorithm"**
   - Submodular functions may be minimized in polynomial time with similar algorithms than for the binary case
   - NB: existing (less efficient) reduction to submodular set-functions defined on a ring family (Schrijver, 2000)

Minimization of submodular functions
Projected subgradient descent

- **For simplicity**: discretizing all sets $\mathcal{X}_i$, $i = 1, \dots, n$ to k elements

- Assume **Lispschitz-continuity**: $\forall x, e_i,\ |H(x + e_i) - H(x)| \leqslant B$

Minimization of submodular functions
Projected subgradient descent

- **For simplicity**: discretizing all sets $\mathcal{X}_i$, $i = 1, \dots, n$ to k elements

- Assume **Lispschitz-continuity**: $\forall x, e_i,\ |H(x + e_i) - H(x)| \leqslant B$

- **Projected subgradient descent**
  - Convergence rate of $O(nkB/\sqrt{t})$ after t iterations
  - Cost of each iteration $O(nk \log(nk))$
  - Reasonable scaling with respect to discretization

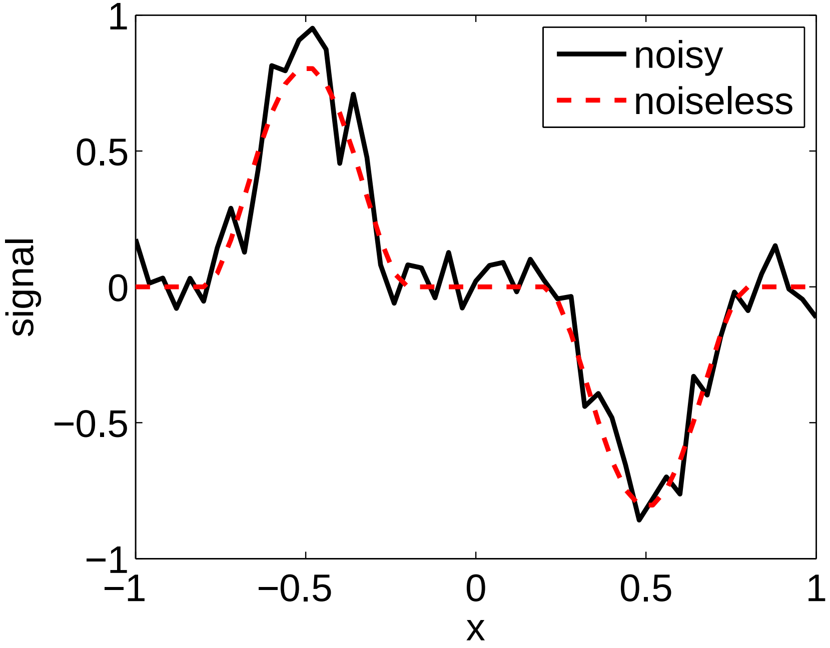
$$\widetilde{O}\Big(\frac{n^3}{\varepsilon^3}\Big) \text{ for continuous domains}$$

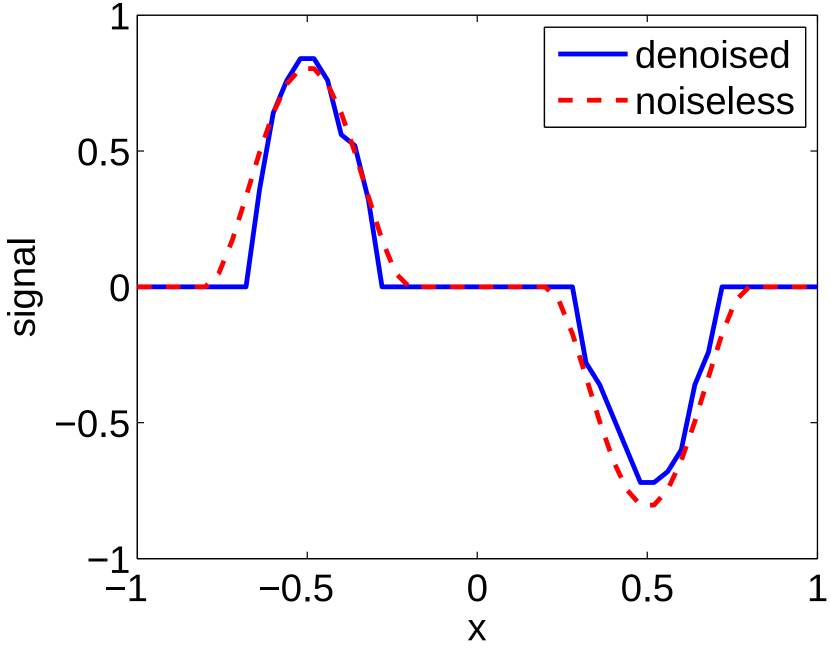
- **Frank-Wolfe / conditional gradient**

Empirical simulations (online code)

- Signal processing example: $H : [-1,1]^n \to \mathbb{R}$ with $\alpha < 1$

$$H(x) = \frac{1}{2}\sum_{i=1}^{n}(x_i - z_i)^2 + \lambda \sum_{i=1}^{n}|x_i|^\alpha + \mu \sum_{i=1}^{n-1}(x_i - x_{i+1})^2$$

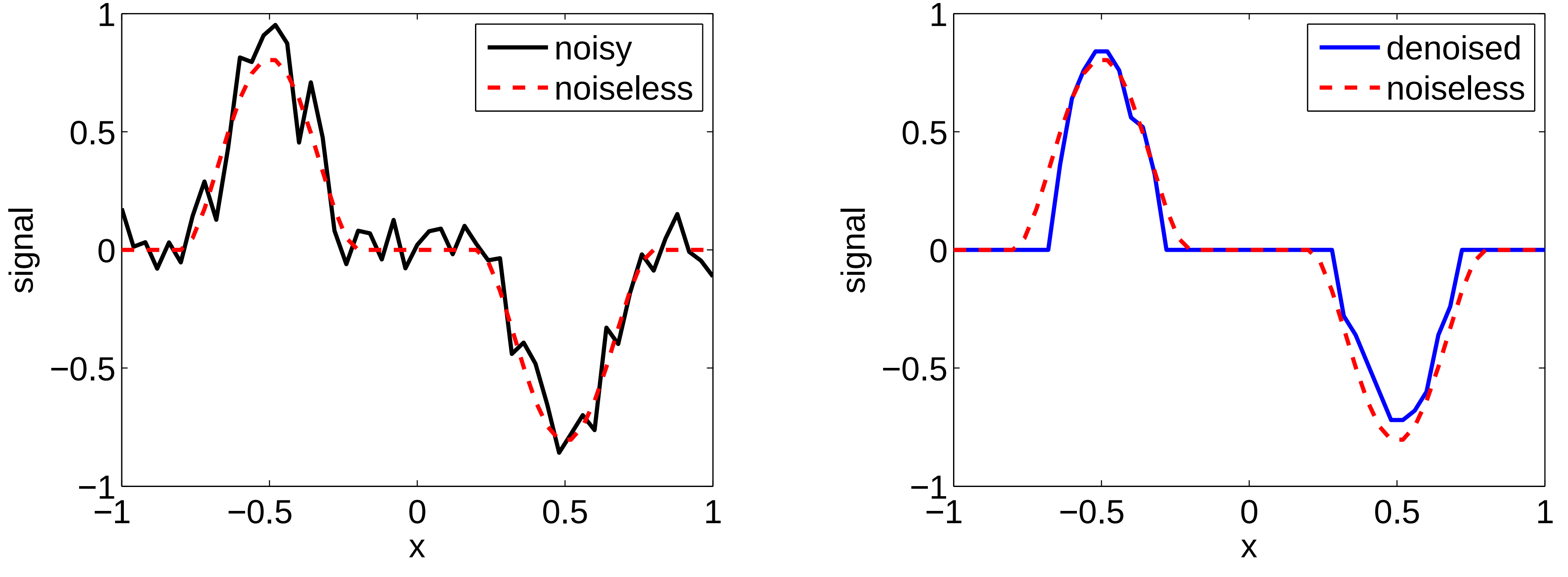




Empirical simulations (online code)

- Signal processing example: $H : [-1,1]^n \to \mathbb{R}$ with $\alpha < 1$

$$H(x) = \frac{1}{2}\sum_{i=1}^{n}(x_i - z_i)^2 + \lambda \sum_{i=1}^{n} |x_i|^\alpha + \mu \sum_{i=1}^{n-1}(x_i - x_{i+1})^2$$

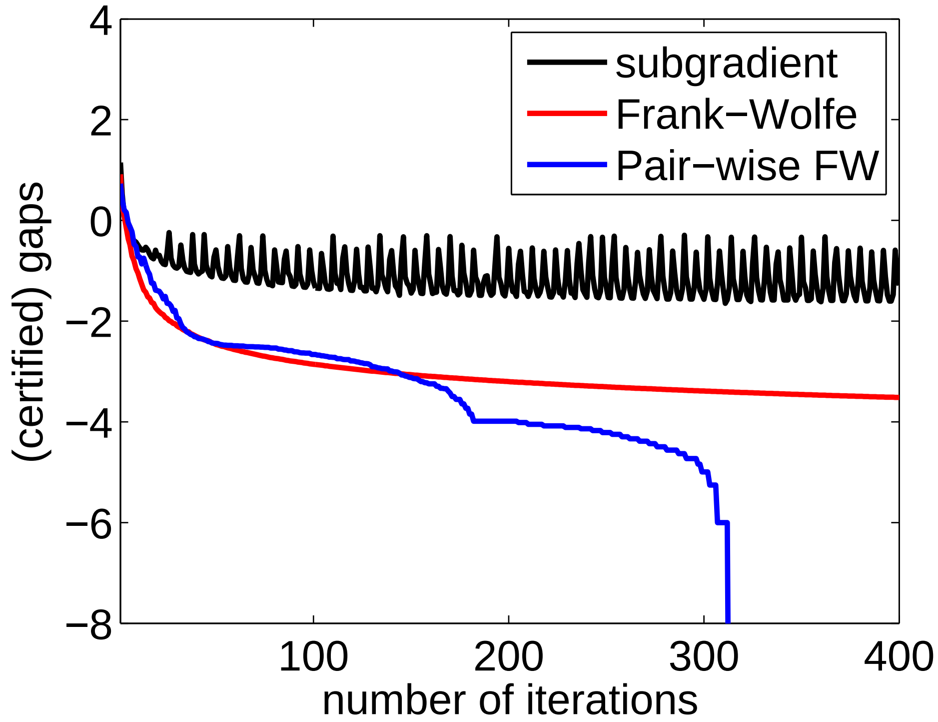


- Generalization to other proximal operators for non-convex regularizers

Empirical simulations

- Signal processing example: $H : [-1,1]^n \to \mathbb{R}$ with $\alpha < 1$

$$H(x) = \frac{1}{2}\sum_{i=1}^{n}(x_i - z_i)^2 + \lambda \sum_{i=1}^{n} |x_i|^\alpha + \mu \sum_{i=1}^{n-1}(x_i - x_{i+1})^2$$

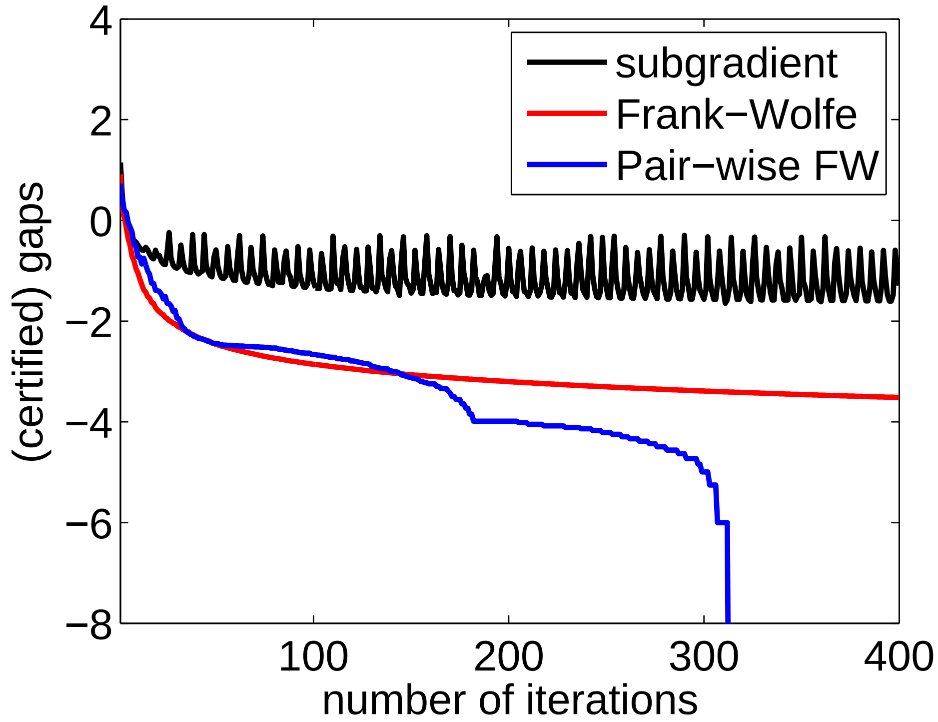


- Pair-wise Frank-Wolfe (Lacoste-Julien and Jaggi, 2015)

Empirical simulations (online code)

- Signal processing example: $H : [-1,1]^n \to \mathbb{R}$ with $\alpha < 1$

$$H(x) = \frac{1}{2}\sum_{i=1}^{n}(x_i - z_i)^2 + \lambda \sum_{i=1}^{n} |x_i|^\alpha + \mu \sum_{i=1}^{n-1}(x_i - x_{i+1})^2$$



- Pair-wise Frank-Wolfe (Lacoste-Julien and Jaggi, 2015)

Conclusion

- **Submodular function and convex optimization**
  - From discrete to continuous domains
  - Extensions to product measures
  - Direct link with one-dimensional multi-marginal optimal transport
  - Application: proximal operator for non-convex regularizers

Conclusion

- **Submodular function and convex optimization**
  - From discrete to continuous domains
  - Extensions to product measures
  - Direct link with one-dimensional multi-marginal optimal transport
  - Application: proximal operator for non-convex regularizers

- **On-going work and extensions**
  - Optimal transport beyond submodular functions
  - Beyond discretization
  - Beyond minimization
  - Sums of simple submodular functions (Jegelka et al., 2013)
  - Mean-field inference in log-supermodular models (Djolonga and Krause, 2015)

Postdoc opportunities in downtown Paris

- **Machine learning group at INRIA - Ecole Normale Supérieure**
  - Two postdoc positions (2 years)
  - One junior researcher position (4 years)

References

F. Bach. *Learning with Submodular Functions: A Convex Optimization Perspective*, volume 6 of *Foundations and Trends in Machine Learning*. NOW, 2013.

F. Bach. Submodular functions: from discrete to continous domains. Technical Report 1511.00394-v2, HAL, 2015.

G. Carlier. On a class of multidimensional optimal transportation problems. *Journal of Convex Analysis*, 10(2):517–530, 2003.

J. Djolonga and A. Krause. Scalable variational inference in log-supermodular models. In *International Conference on Machine Learning (ICML)*, 2015.

S. Fujishige. *Submodular Functions and Optimization*. Elsevier, 2005.

S. Jegelka, F. Bach, and S. Sra. Reflection methods for user-friendly submodular optimization. In *Advances in Neural Information Processing Systems (NIPS)*, 2013.

S. Kim and M. Kojima. Exact solutions of some nonconvex quadratic optimization problems via SDP and SOCP relaxations. *Computational Optimization and Applications*, 26(2):143–154, 2003.

S. Lacoste-Julien and M. Jaggi. On the global linear convergence of frank-wolfe optimization variants. In *Advances in Neural Information Processing Systems (NIPS)*, 2015.

G. G. Lorentz. An inequality for rearrangements. *American Mathematical Monthly*, 60(3):176–179, 1953.

P. Milgrom and C. Shannon. Monotone comparative statics. *Econometrica: Journal of the Econometric Society*, pages 157–180, 1994.

F. Santambrogio. *Optimal Transport for Applied Mathematicians*. Springer, 2015.

A. Schrijver. A combinatorial algorithm minimizing submodular functions in strongly polynomial time. *Journal of Combinatorial Theory, Series B*, 80(2):346–355, 2000.

D. M. Topkis. Minimizing a submodular function on a lattice. *Operations research*, 26(2):305–321, 1978.

C. Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.