On Polynomial-Time Solvability of Combinatorial Markov Random Fields

Shaoning Han

Department of Industrial & Systems Engineering
University of Southern California

IOS Conference, March 2024

Collaborators

Andres Gomez
ISE, USC

Jong-Shi Pang
ISE, USC

Markov random field

Markov random field An MRF model is defined on an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where

- random variable $X_i = x_i + \epsilon_i$ with $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ for $i \in \mathcal{V}$
- X_i is only dependent on its neighbors and independent of others

MRF inference Infer true values of $\{X_i\}_{i \in \mathcal{V}}$ from noisy observations $\{a_i\}_{i \in \mathcal{V}}$

$$\underset{\boldsymbol{x} \in \mathbb{R}^n}{\textbf{minimize}} \ \sum_{i \in \mathcal{V}} \underbrace{\frac{1}{\sigma_i^2}(x_i - a_i)^2}_{\text{fitness}} + \sum_{(i,j) \in \mathcal{E}} \underbrace{\frac{1}{\sigma_{ij}}(x_i - x_j)^2}_{\text{smoothness}}$$

$$\textbf{subject to} \ \ell_i \le x_i \le u_i \ \forall i \in \mathcal{V}$$

Markov random field

Markov random field An MRF model is defined on an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where

- random variable $X_i = x_i + \epsilon_i$ with $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ for $i \in \mathcal{V}$
- X_i is only dependent on its neighbors and independent of others

MRF inference Infer true values of $\{X_i\}_{i \in \mathcal{V}}$ from noisy observations $\{a_i\}_{i \in \mathcal{V}}$

$$\underset{\boldsymbol{x} \in \mathbb{R}^n}{\textbf{minimize}} \sum_{i \in \mathcal{V}} \underbrace{\frac{1}{\sigma_i^2}(x_i - a_i)^2}_{\text{fitness}} + \sum_{(i,j) \in \mathcal{E}} \underbrace{\frac{1}{\sigma_{ij}}(x_i - x_j)^2}_{\text{smoothness}}$$

$$\textbf{subject to } \ell_i \le x_i \le u_i \ \forall i \in \mathcal{V}$$

Markov random field

Markov random field An MRF model is defined on an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where

- random variable $X_i = x_i + \epsilon_i$ with $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ for $i \in \mathcal{V}$
- X_i is only dependent on its neighbors and independent of others

MRF inference Infer true values of $\{X_i\}_{i\in\mathcal{V}}$ from noisy observations $\{a_i\}_{i\in\mathcal{V}}$

$$\begin{aligned} &\underset{\boldsymbol{x}\in\mathbb{R}^n}{\textbf{minimize}} \; \sum_{i\in\mathcal{V}} \underbrace{\frac{1}{\sigma_i^2}(x_i - a_i)^2}_{\text{fitness}} + \sum_{(i,j)\in\mathcal{E}} \underbrace{\frac{1}{\sigma_{ij}}(x_i - x_j)^2}_{\text{smoothness}} \\ &\textbf{subject to } \ell_i \le x_i \le u_i \; \forall i \in \mathcal{V} \end{aligned}$$

- x_i: true values (decision variables)
- σ_{ij}: correlation coefficients
- $\ell_i \in \mathbb{R} \cup \{-\infty\}$, $u_i \in \mathbb{R} \cup \{+\infty\}$

Markov random field

Markov random field An MRF model is defined on an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where

- random variable $X_i = x_i + \epsilon_i$ with $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ for $i \in \mathcal{V}$
- X_i is only dependent on its neighbors and independent of others

MRF inference Infer true values of $\{X_i\}_{i \in \mathcal{V}}$ from noisy observations $\{a_i\}_{i \in \mathcal{V}}$

$$\underset{\boldsymbol{x} \in \mathbb{R}^n}{\textbf{minimize}} \sum_{i \in \mathcal{V}} \underbrace{\frac{1}{\sigma_i^2}(x_i - a_i)^2}_{\text{fitness}} + \sum_{(i,j) \in \mathcal{E}} \underbrace{\frac{1}{\sigma_{ij}}(x_i - x_j)^2}_{\text{smoothness}}$$

$$\textbf{subject to } \ell_i \leq x_i \leq u_i \ \forall i \in \mathcal{V}$$

- Markov property
- Negative correlation
- Smoothness / pairwise similarity

Application

1D MRF

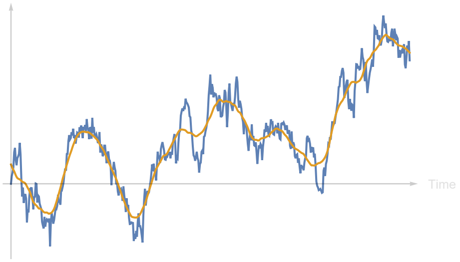
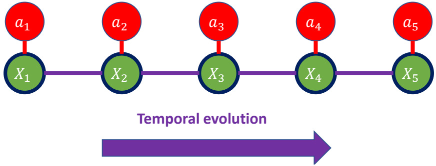
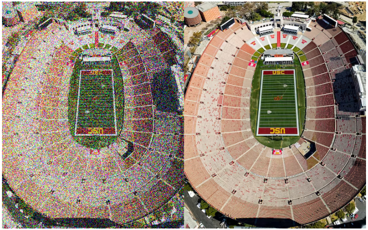


Figure: Weiner Process - Time Series

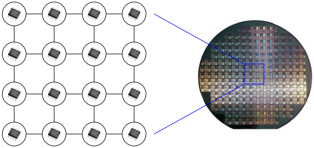


Application

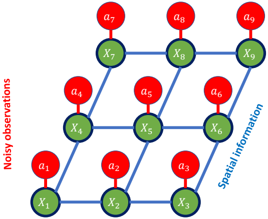
2D MRF



(a) Image denoising

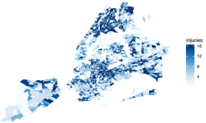


(b) Manufacturing (Schrunner et al. 2017)

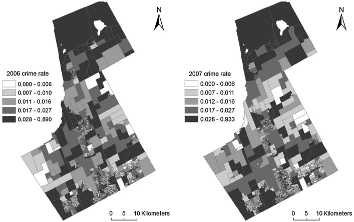


Application

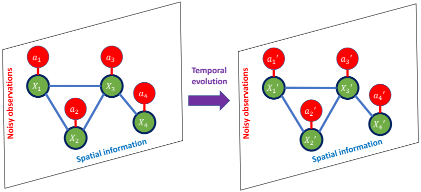
3D MRF



(a) Epidemiology (Morris et al. 2019)



(b) Criminology (Law et al. 2014)



Sparse MRF inference

Assumption the underlying statistical process is sparse

$$\underset{x\in\mathbb{R}^n}{\textbf{minimize}} \sum_{i\in\mathcal{V}} \frac{1}{\sigma_i^2} \underbrace{(x_i - a_i)^2}_{\text{fitness}} + \sum_{(i,j)\in\mathcal{E}} \frac{1}{\sigma_{ij}^2} \underbrace{(x_i - x_j)^2}_{\text{smoothness}} + \underbrace{\lambda \|\boldsymbol{x}\|_0}_{\text{sparsity}}$$

$$\textbf{subject to } \ell_i \le x_i \le u_i \ \forall i \in \mathcal{V}$$

Sparse MRF inference

Assumption the underlying statistical process is sparse

$$\underset{x\in\mathbb{R}^n}{\textbf{minimize}} \sum_{i\in\mathcal{V}} \frac{1}{\sigma_i^2} \underbrace{(x_i - a_i)^2}_{\text{fitness}} + \sum_{(i,j)\in\mathcal{E}} \frac{1}{\sigma_{ij}^2} \underbrace{(x_i - x_j)^2}_{\text{smoothness}} + \lambda \underbrace{\sum_{i\in\mathcal{V}} z_i}_{\text{sparsity}}$$

$$\textbf{subject to } \ell_i z_i \leq x_i \leq u_i z_i \ \forall i \in \mathcal{V}$$
$$z \in \{0,1\}^n$$

Define $0 \cdot (\pm\infty) = 0$

- z_i indicates if x_i is zero: $[z_i = 0 \Rightarrow x_i = 0]$ & $[z_i = 1 \Rightarrow \ell_i \leq x_i \leq u_i]$
- If $\ell_i = -\infty$ and $u_i = +\infty$, then $\ell_i z_i \leq x_i \leq u_i z_i \Leftrightarrow x_i(1 - z_i) = 0$

Robust MRF

Assumption a few observations a_i are corrupted by gross outliers

$$\underset{\mathcal{U}\subseteq\mathcal{V},x\in\mathbb{R}^n}{\textbf{minimize}} \sum_{i\in\mathcal{V}\setminus\mathcal{U}} \frac{1}{\sigma_i^2}\underbrace{(x_i-a_i)^2}_{\text{fitness}} + \sum_{(i,j)\in\mathcal{E}} \frac{1}{\sigma_{ij}^2}\underbrace{(x_i-x_j)^2}_{\text{smoothness}} + \underbrace{\lambda|\mathcal{U}|}_{\text{robustness}}$$

$$\textbf{subject to } \ell_i \le x_i \le u_i \ \forall i \in \mathcal{V}$$

- $\mathcal{U}$: the set of outliers

Introducing binary variables $[z_i = 1 \Leftrightarrow i \in \mathcal{U}] \Rightarrow$ a MIP formulation

Robust MRF inference

Assumption a few observations a_i are corrupted by gross outliers

$$\underset{z,w,x\in\mathbb{R}^n}{\textbf{minimize}} \sum_{i\in\mathcal{V}} \frac{1}{\sigma_i^2} \underbrace{(x_i - a_i - w_i)^2}_{\text{fitness}} + \sum_{(i,j)\in\mathcal{E}} \frac{1}{\sigma_{ij}^2} \underbrace{(x_i - x_j)^2}_{\text{smoothness}} + \lambda \underbrace{\sum_{i\in\mathcal{V}} z_i}_{\text{robustness}}$$

$$\textbf{subject to } \ell_i \leq x_i \leq u_i \ \forall i \in \mathcal{V}$$

$$w_i(1 - z_i) = 0 \ \forall i \in \mathcal{V} \ , \ z \in \{0, 1\}^n$$

Equivalence

- $z_i = 0$: $w_i = 0 \Rightarrow a_i$ is not an outlier
- $z_i = 1$: $w_i = a_i - x_i$ at the optimal solution $\Rightarrow a_i$ is an outlier

Robust MRF inference

Assumption a few observations a_i are corrupted by gross outliers

$$\underset{z,w,x\in\mathbb{R}^n}{\textbf{minimize}} \sum_{i\in\mathcal{V}} \frac{1}{\sigma_i^2} \underbrace{(x_i - a_i - w_i)^2}_{\text{fitness}} + \sum_{(i,j)\in\mathcal{E}} \frac{1}{\sigma_{ij}^2} \underbrace{(x_i - x_j)^2}_{\text{smoothness}} + \lambda \underbrace{\sum_{i\in\mathcal{V}} z_i}_{\text{robustness}}$$

$$\textbf{subject to } \ell_i \le x_i \le u_i \ \forall i \in \mathcal{V}$$

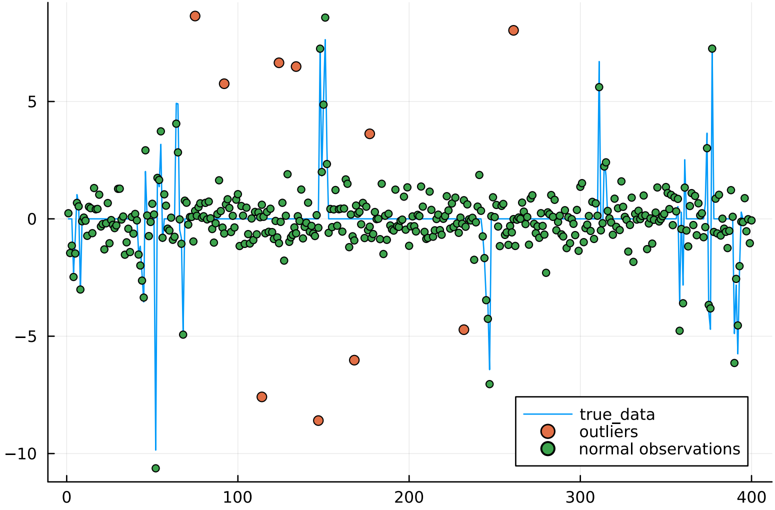
$$\tilde{\ell} z_i \le w_i \le \tilde{u} z_i, \ z \in \{0,1\}^n$$

Equivalence

- $z_i = 0$: $w_i = 0 \Rightarrow a_i$ is not an outlier
- $z_i = 1$: $w_i = a_i - x_i$ at the optimal solution $\Rightarrow a_i$ is an outlier
- $\tilde{\ell} = -\infty$ and $\tilde{u} = +\infty$

1D Example – combinatorial MRF

- Smooth
- Outliers



Our contribution

Sparse and robust MRF can be put in the form

$$\underset{x\in\mathbb{R}^n, z\in\{0,1\}^n}{\textbf{minimize}} \left\{\frac{1}{2}x^\top Qx + b^\top x + c^\top z : \ \ell_i z_i \leq x_i \leq u_i z_i \ \forall i\right\},$$

where $Q \succeq 0$

- In general, the problem is NP-hard, e.g., if $f(x) =$ the obj of OLS

Our contribution

Sparse and robust MRF can be put in the form

$$\underset{x\in\mathbb{R}^n, z\in\{0,1\}^n}{\textbf{minimize}} \left\{\frac{1}{2}x^\top Qx + b^\top x + c^\top z :\ \ell_i z_i \leq x_i \leq u_i z_i \ \forall i\right\},$$

where $Q \succeq 0$ and $Q_{ij} \leq 0 \ \forall i \neq j$

Theorem (Polynomial solvability)

The problem of sparse/robust MRF can be solved as a binary submodular minimization problem and thus is **(strongly) polynomially** *solvable.*

Solution path / hyperparameter selection

Question When the sparsity pattern (or the number of outliers) are unknown, how to choose λ in

$$\operatorname*{\mathbf{minimum}}_{\ell \leq x \leq u} \frac{1}{2} x^\top Q x + b^\top x + \lambda \|x\|_0 \tag{$\star$}$$

Solution path / hyperparameter selection

Question When the sparsity pattern (or the number of outliers) are unknown, how to choose λ in

$$p(\lambda) \overset{\text{def}}{=} \underset{\ell \le x \le u}{\textbf{minimum}} \, \frac{1}{2} x^\top Q x + b^\top x + \lambda \|x\|_0 \tag{$\star$}$$

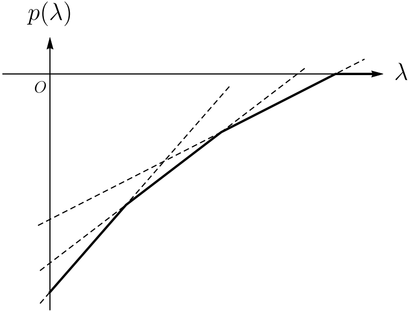
Answer Compute <u>all possible</u> $p(\lambda)$ and choose a desired one! (AIC, etc.)

Solution path / hyperparameter selection

Question When the sparsity pattern (or the number of outliers) are unknown, how to choose λ in

$$p(\lambda) \stackrel{\text{def}}{=} \underset{\ell \le x \le u}{\textbf{minimum}} \ \frac{1}{2} x^\top Q x + b^\top x + \lambda \|x\|_0 \qquad (\star)$$

Answer Compute all possible $p(\lambda)$ and choose a desired one! (AIC, etc.)



Solution path / hyperparameter selection

Question When the sparsity pattern (or the number of outliers) are unknown, how to choose λ in

$$p(\lambda) \stackrel{\text{def}}{=} \operatorname*{\mathbf{minimum}}_{\ell \le x \le u} \frac{1}{2} x^\top Q x + b^\top x + \lambda \|x\|_0 \qquad (\star)$$

Answer Compute <u>all possible</u> $p(\lambda)$ and choose a desired one! (AIC, etc.)

Proposition

Solution path $p(\bullet)$ is a concave increasing piecewise affine function of λ, which consists of at most $n + 1$ pieces. Moreover, it can be computed in polynomial time.

<u>Free of hyper-parameter tuning!</u>

Experimental results - robust MRF

Submodular v.s. MIP

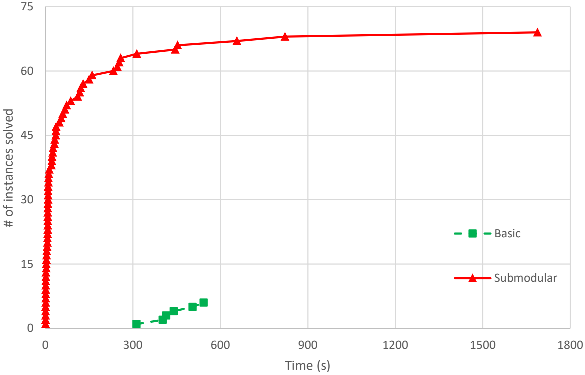


Figure: Number of instances solved as a function of time

- Solvability: Submodular 92% versus MIP 8%
- Solution time: 700x speed-up!

A touch of math

We will

- show sparse/robust MRF is **theoretically tractable** by reducing it into a binary submodular minimization problem

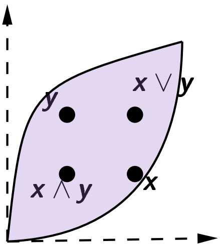
- make sparse/robust MRF **practically tractable** by designing a parametric pivoting method to efficiently compute extremal bases

Lattices

Meet and Join Given $x, y \in \mathbb{R}^n$, define

- Meet: $x \wedge y \stackrel{\text{def}}{=} (\min\{x_i, y_i\})_i$
- Join: $x \vee y \stackrel{\text{def}}{=} (\max\{x_i, y_i\})_i$

Lattice A set $\mathcal{L} \subset \mathbb{R}^n$ is a lattice if $[x, y \in \mathcal{L} \Rightarrow x \vee y, x \wedge y \in \mathcal{L}]$

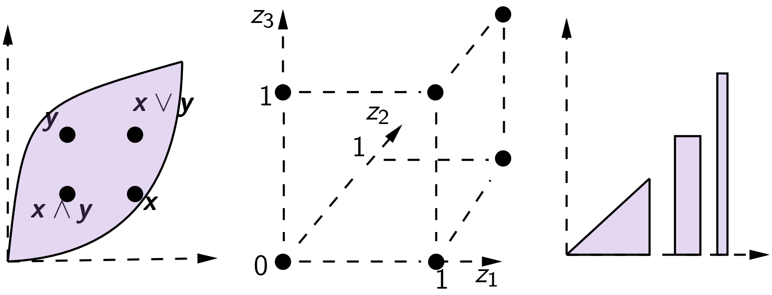


Lattices

Meet and Join Given $x, y \in \mathbb{R}^n$, define

- Meet: $x \wedge y \stackrel{\text{def}}{=} (\min\{x_i, y_i\})_i$
- Join: $x \vee y \stackrel{\text{def}}{=} (\max\{x_i, y_i\})_i$

Lattice A set $\mathcal{L} \subset \mathbb{R}^n$ is a lattice if $[x, y \in \mathcal{L} \Rightarrow x \vee y, x \wedge y \in \mathcal{L}]$



Submodularity

Submodularity Given a lattice $\mathcal{L} \subseteq \mathbb{R}^n$, a function $f : \mathcal{L} \to \mathbb{R}$ is submodular if

$$f(x) + f(y) \geq f(x \wedge y) + f(x \vee y) \; \forall x, y \in \mathcal{L}$$

Remarks

- If $\mathcal{L} \subseteq \{0, 1\}^n$, then f is a binary/set submodular function
- If $f \in \mathcal{C}^2(\mathbb{R}^n)$, submodularity over $\mathbb{R}^n \Leftrightarrow \frac{\partial^2 f}{\partial y_i \partial y_j}(y) \leq 0 \; \forall i \neq j$

Submodularity

Submodularity Given a lattice $\mathcal{L} \subseteq \mathbb{R}^n$, a function $f : \mathcal{L} \to \mathbb{R}$ is submodular if

$$f(x) + f(y) \geq f(x \wedge y) + f(x \vee y) \; \forall x, y \in \mathcal{L}$$

Remarks

- If $\mathcal{L} \subseteq \{0,1\}^n$, then f is a binary/set submodular function
- If $f \in \mathcal{C}^2(\mathbb{R}^n)$, submodularity over $\mathbb{R}^n \Leftrightarrow \frac{\partial^2 f}{\partial y_i \partial y_j}(y) \leq 0 \; \forall i \neq j$

Examples

- $n = 1 \Rightarrow f(x)$ is submodular
- $f(x) = c^\top x$ is submodular
- $f(x) = x^\top \begin{bmatrix} 5 & -1 & -3 \\ -1 & 3 & -2 \\ -3 & -2 & 7 \end{bmatrix} x$ is submodular ($Q_{ij} \leq 0$ for all $i \neq j$)

Submodularity

Submodularity Given a lattice $\mathcal{L} \subseteq \mathbb{R}^n$, a function $f : \mathcal{L} \to \mathbb{R}$ is submodular if

$$f(x) + f(y) \geq f(x \wedge y) + f(x \vee y) \; \forall x, y \in \mathcal{L}$$

Remarks

- If $\mathcal{L} \subseteq \{0,1\}^n$, then f is a binary/set submodular function
- If $f \in \mathcal{C}^2(\mathbb{R}^n)$, submodularity over $\mathbb{R}^n \Leftrightarrow \frac{\partial^2 f}{\partial y_i \partial y_j}(y) \leq 0 \; \forall i \neq j$

Key observation

Lemma (Topkis (1978))

Given a lattice $\mathcal{L} \in \mathbb{R}^m \times \mathbb{R}^n$ and a submodular function $f : \mathcal{L} \to \mathbb{R}$, the marginal function

$$v(z) \stackrel{def}{=} \underset{x \in \mathbb{R}^m}{\textbf{minimum}} \{f(x, z) : (x, z) \in \mathcal{L}\}$$

is submodular on the lattice $\mathrm{proj}_z \stackrel{def}{=} \{z : \exists x \text{ s.t. } (x, z) \in \mathcal{L}\}$.

Nonnegative case

Assume $\ell \in \mathbb{R}_+^n$ and get back to

$$\underset{x \in \mathbb{R}^n, z \in \{0,1\}^n}{\textbf{minimize}} \left\{ \frac{1}{2} x^\top Q x + b^\top x + c^\top z : \ \ell_i z_i \leq x_i \leq u_i z_i \ \forall i \right\} \qquad (\star)$$

Objective

$$f(x) + c^\top z := \frac{1}{2} x^\top Q x + b^\top x + c^\top z$$

Feasible region $\prod_{i \in \mathcal{V}} \{(x_i, z_i) \in \mathbb{R} \times \{0, 1\} : \ell_i z_i \leq x_i \leq z_i u_i\}$

Nonnegative case

Assume $\ell \in \mathbb{R}^n_+$ and get back to

$$\operatorname*{\mathbf{minimize}}_{x \in \mathbb{R}^n, z \in \{0,1\}^n} \left\{ \frac{1}{2} x^\top Q x + b^\top x + c^\top z : \; \ell_i z_i \le x_i \le u_i z_i \; \forall i \right\} \qquad (\star)$$

Objective is submodular due to $Q_{ij} \le 0 \; \forall i \ne j$

$$f(x) + c^\top z := \frac{1}{2} x^\top Q x + b^\top x + c^\top z$$

Feasible region $\prod_{i \in \mathcal{V}} \{(x_i, z_i) \in \mathbb{R} \times \{0,1\} : \ell_i z_i \le x_i \le z_i u_i\}$

Nonnegative case

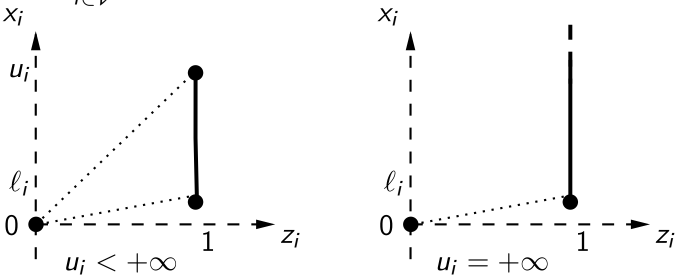
Assume $\ell \in \mathbb{R}^n_+$ and get back to

$$\underset{x\in\mathbb{R}^n, z\in\{0,1\}^n}{\textbf{minimize}} \left\{ \frac{1}{2}x^\top Qx + b^\top x + c^\top z : \ \ell_i z_i \le x_i \le u_i z_i \ \forall i \right\} \qquad (\star)$$

Objective is submodular due to $Q_{ij} \le 0 \ \forall i \ne j$

$$f(x) + c^\top z := \frac{1}{2}x^\top Qx + b^\top x + c^\top z$$

Feasible region $\prod_{i\in\mathcal{V}} \{(x_i, z_i) \in \mathbb{R} \times \{0,1\} : \ell_i z_i \le x_i \le z_i u_i\}$



Nonnegative case

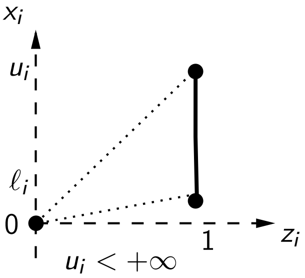
Assume $\ell \in \mathbb{R}^n_+$ and get back to

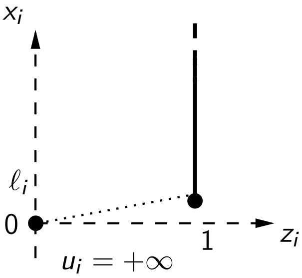
$$\underset{x \in \mathbb{R}^n, z \in \{0,1\}^n}{\textbf{minimize}} \left\{ \frac{1}{2} x^\top Q x + b^\top x + c^\top z : \ \ell_i z_i \leq x_i \leq u_i z_i \ \forall i \right\} \qquad (\star)$$

Objective is submodular due to $Q_{ij} \leq 0 \ \forall i \neq j$

$$f(x) + c^\top z := \frac{1}{2} x^\top Q x + b^\top x + c^\top z$$

Feasible region is a lattice due to $\ell \geq 0$





Nonnegative case

Assume $\ell \in \mathbb{R}^n_+$ and get back to

$$\underset{x\in\mathbb{R}^n, z\in\{0,1\}^n}{\textbf{minimize}} \left\{ \frac{1}{2}x^\top Qx + b^\top x + c^\top z : \ \ell_i z_i \le x_i \le u_i z_i \ \forall i \right\} \qquad (\star)$$

Objective is submodular due to $Q_{ij} \le 0 \ \forall i \ne j$

$$f(x) + c^\top z := \frac{1}{2}x^\top Qx + b^\top x + c^\top z$$

Feasible region is a lattice due to $\ell \ge 0$

Thus,

$$(\star) \Leftrightarrow \underset{\boldsymbol{z}\in\{0,1\}^n}{\textbf{minimize}}\, v(\boldsymbol{z}) + \boldsymbol{c}^\top \boldsymbol{z}$$

where $v(\boldsymbol{z}) = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\textbf{minimum}} \{ f(\boldsymbol{x}) : \ \boldsymbol{\ell} \circ \boldsymbol{z} \le \boldsymbol{x} \le \boldsymbol{u} \circ \boldsymbol{z} \}$ is a binary submodular function and can be evaluated by solving a convex program

General case

Assume $\ell \notin \mathbb{R}_+^n$,

$$\underset{x\in\mathbb{R}^n, z\in\{0,1\}^n}{\textbf{minimize}} \left\{\frac{1}{2}x^\top Q x + b^\top x + c^\top z : \ \ell_i z_i \le x_i \le u_i z_i \ \forall i\right\} \qquad (\star)$$

Issue the feasible region is not a lattice if $\ell_i < 0$

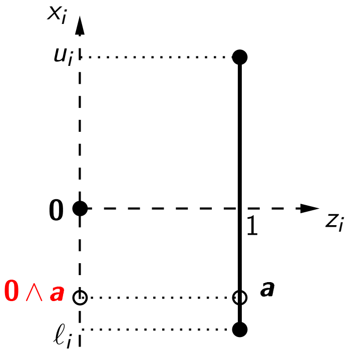


Figure: Region of $\{(x_i, z_i) \in \mathbb{R} \times \{0,1\} : \ell_i z_i \le x_i \le z_i u_i\}$

General case

Assume $\ell \notin \mathbb{R}^n_+$,

$$\underset{x \in \mathbb{R}^n, z \in \{0,1\}^n}{\textbf{minimize}} \left\{ \frac{1}{2} x^\top Q x + b^\top x + c^\top z : \ \ell_i z_i \leq x_i \leq u_i z_i \ \forall i \right\} \qquad (\star)$$

Issue the feasible region is not a lattice if $\ell_i < 0$

Idea If $\ell_i < 0$ and $u_i > 0$, then

$$\ell_i z_i \leq x_i \leq u_i z_i, \ z_i \in \{0,1\} \Leftrightarrow \begin{cases} z_i = z_i^+ + (1 - z_i^-) \\ z_i^+ \in \{0,1\}, \ z_i^- \in \{0,1\} \\ \ell_i(1 - z_i^-) \leq x_i \leq u_i z_i^+ \\ z_i^- \geq z_i^+ \end{cases}$$

- Split z_i into two parts z_i^+ and $(1 - z_i^-)$

General case

Assume $\ell \notin \mathbb{R}^n_+$,

$$\underset{x\in\mathbb{R}^n, z\in\{0,1\}^n}{\textbf{minimize}} \left\{\frac{1}{2}x^\top Qx + b^\top x + c^\top z :\ \ell_i z_i \le x_i \le u_i z_i\ \forall i\right\} \qquad (\star)$$

Issue the feasible region is not a lattice if $\ell_i < 0$

Idea If $\ell_i < 0$ and $u_i > 0$, then

$$\ell_i z_i \le x_i \le u_i z_i,\ z_i \in \{0,1\} \Leftrightarrow \begin{cases} z_i = z_i^+ + (1 - z_i^-) \\ z_i^+ \in \{0,1\},\ z_i^- \in \{0,1\} \\ \ell_i(1 - z_i^-) \le x_i \le u_i z_i^+ \\ z_i^- \ge z_i^+ \end{cases}$$

- Split z_i into two parts z_i^+ and $(1 - z_i^-)$
- $z_i^+ = 0 \Rightarrow [x_i]_+ \overset{\text{def}}{=} \max\{x_i, 0\} = 0 \Leftrightarrow x_i \le 0$

General case

Assume $\ell \notin \mathbb{R}^n_+$,

$$\underset{x\in\mathbb{R}^n, z\in\{0,1\}^n}{\textbf{minimize}} \left\{ \frac{1}{2} x^\top Q x + b^\top x + c^\top z : \ \ell_i z_i \le x_i \le u_i z_i \ \forall i \right\} \qquad (\star)$$

Issue the feasible region is not a lattice if $\ell_i < 0$

Idea If $\ell_i < 0$ and $u_i > 0$, then

$$\ell_i z_i \le x_i \le u_i z_i, \ z_i \in \{0,1\} \Leftrightarrow \begin{cases} z_i = z_i^+ + (1 - z_i^-) \\ z_i^+ \in \{0,1\}, \ z_i^- \in \{0,1\} \\ \ell_i(1 - z_i^-) \le x_i \le u_i z_i^+ \\ z_i^- \ge z_i^+ \end{cases}$$

- Split z_i into two parts z_i^+ and $(1 - z_i^-)$
- $z_i^+ = 0 \Rightarrow [x_i]_+ \overset{\text{def}}{=} \max\{x_i, 0\} = 0 \Leftrightarrow x_i \le 0$
- $1 - z_i^- = 0 \Rightarrow [x_i]_- \overset{\text{def}}{=} \max\{-x_i, 0\} = 0 \Leftrightarrow x_i \ge 0$

General case

Assume $\ell \notin \mathbb{R}^n_+$,

$$\underset{x\in\mathbb{R}^n, z\in\{0,1\}^n}{\textbf{minimize}} \left\{\frac{1}{2}x^\top Qx + b^\top x + c^\top z : \ \ell_i z_i \le x_i \le u_i z_i \ \forall i\right\} \qquad (\star)$$

Issue the feasible region is not a lattice if $\ell_i < 0$

Idea If $\ell_i < 0$ and $u_i > 0$, then

$$\ell_i z_i \le x_i \le u_i z_i, \ z_i \in \{0,1\} \Leftrightarrow \begin{cases} z_i = z_i^+ + (1 - z_i^-) \\ z_i^+ \in \{0,1\}, \ z_i^- \in \{0,1\} \\ \ell_i(1 - z_i^-) \le x_i \le u_i z_i^+ \\ z_i^- \ge z_i^+ \end{cases}$$

- Split z_i into two parts z_i^+ and $(1 - z_i^-)$
- $z_i^+ = 0 \Rightarrow [x_i]_+ \overset{\text{def}}{=} \max\{x_i, 0\} = 0 \Leftrightarrow x_i \le 0$
- $1 - z_i^- = 0 \Rightarrow [x_i]_- \overset{\text{def}}{=} \max\{-x_i, 0\} = 0 \Leftrightarrow x_i \ge 0$
- $[x_i]_+$ and $[x_i]_-$ can not be both nonzero $\Rightarrow z_i^+ + (1 - z_i^-) \le 1$

General case

Assume $\ell \notin \mathbb{R}^n_+$,

$$\underset{x\in\mathbb{R}^n, z\in\{0,1\}^n}{\textbf{minimize}} \left\{\frac{1}{2}x^\top Qx + b^\top x + c^\top z : \ \ell_i z_i \le x_i \le u_i z_i \ \forall i\right\} \qquad (\star)$$

Issue the feasible region is not a lattice if $\ell_i < 0$

Idea If $\ell_i < 0$ and $u_i > 0$, then

$$\ell_i z_i \le x_i \le u_i z_i, \ z_i \in \{0,1\} \Leftrightarrow \begin{cases} z_i = z_i^+ + (1 - z_i^-) \\ z_i^+ \in \{0,1\}, \ z_i^- \in \{0,1\} \\ \ell_i(1 - z_i^-) \le x_i \le u_i z_i^+ \\ z_i^- \ge z_i^+ \end{cases} \Rightarrow \text{lattice}$$

- Split z_i into two parts z_i^+ and $(1 - z_i^-)$
- $z_i^+ = 0 \Rightarrow [x_i]_+ \overset{\text{def}}{=} \max\{x_i, 0\} = 0 \Leftrightarrow x_i \le 0$
- $1 - z_i^- = 0 \Rightarrow [x_i]_- \overset{\text{def}}{=} \max\{-x_i, 0\} = 0 \Leftrightarrow x_i \ge 0$
- $[x_i]_+$ and $[x_i]_-$ can not be both nonzero $\Rightarrow z_i^+ + (1 - z_i^-) \le 1$

General case

For simplicity, assume $\ell < \mathbf{0} < u$. Substituting out z_i,

$$\begin{aligned} \underset{\boldsymbol{x},\boldsymbol{z}^+,\boldsymbol{z}^- \in \mathbb{R}^n}{\textbf{minimize}} \ & \frac{1}{2}x^\top Q x + b^\top x + c^\top (z^+ + \mathbf{1} - z^-) \\ \textbf{subject to} \ & \ell_i (1 - z_i^-) \leq x_i \leq u_i z_i^+ \ \forall i \\ & z^- \geq z^+, \ z^+, z^- \in \{0,1\}^n \end{aligned} \qquad (\star)$$

General case

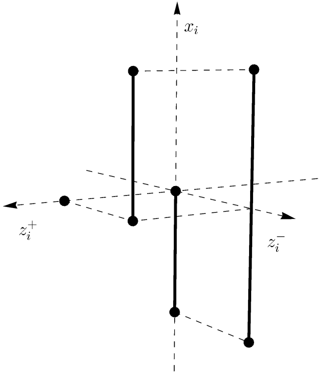
For simplicity, assume $\ell < \mathbf{0} < u$. Substituting out z_i,

$$\begin{aligned} \underset{x, z^+, z^- \in \mathbb{R}^n}{\textbf{minimize}} \ & \frac{1}{2}x^\top Q x + b^\top x + c^\top (z^+ + \mathbf{1} - z^-) \\ \textbf{subject to} \ & \ell_i (1 - z_i^-) \leq x_i \leq u_i z_i^+ \ \forall i \\ & \cancel{z^- \geq z^+}, \ z^+, z^- \in \{0,1\}^n \end{aligned} \tag{$\star$}$$

General case

For simplicity, assume $\ell < \mathbf{0} < u$. Substituting out z_i,

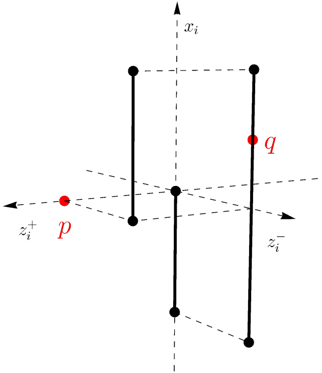
$$\begin{aligned} \underset{x, z^+, z^- \in \mathbb{R}^n}{\textbf{minimize}} \ & \frac{1}{2} x^\top Q x + b^\top x + c^\top (z^+ + \mathbf{1} - z^-) \\ \textbf{subject to} \ & \ell_i (1 - z_i^-) \le x_i \le u_i z_i^+ \ \forall i \\ & \cancel{z^- \ge z^+}, \ z^+, z^- \in \{0,1\}^n \end{aligned} \tag{$\star$}$$



General case

For simplicity, assume $\ell < \mathbf{0} < u$. Substituting out z_i,

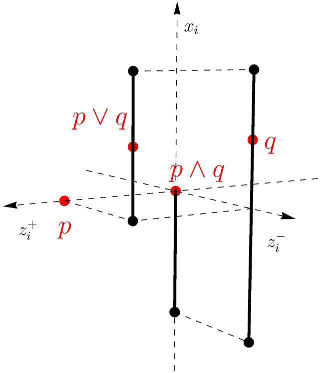
$$\begin{aligned} \underset{x, z^+, z^- \in \mathbb{R}^n}{\textbf{minimize}} \quad & \frac{1}{2} x^\top Q x + b^\top x + c^\top (z^+ + \mathbf{1} - z^-) \\ \textbf{subject to} \quad & \ell_i (1 - z_i^-) \le x_i \le u_i z_i^+ \ \forall i \\ & \cancel{z^- \ge z^+},\ z^+, z^- \in \{0, 1\}^n \end{aligned} \tag{$\star$}$$



General case

For simplicity, assume $\ell < \mathbf{0} < u$. Substituting out z_i,

$$\begin{aligned} \underset{x, z^+, z^- \in \mathbb{R}^n}{\textbf{minimize}} \ & \frac{1}{2} x^\top Q x + b^\top x + c^\top (z^+ + \mathbf{1} - z^-) \\ \textbf{subject to} \ & \ell_i (1 - z_i^-) \le x_i \le u_i z_i^+ \ \forall i \\ & \cancel{z^- \ge z^+}, \ z^+, z^- \in \{0,1\}^n \end{aligned} \tag{$\star$}$$



General case

For simplicity, assume $\ell < \mathbf{0} < u$. Substituting out z_i,

$$\begin{aligned} \underset{\boldsymbol{x}, \boldsymbol{z}^+, \boldsymbol{z}^- \in \mathbb{R}^n}{\textbf{minimize}} \; & \frac{1}{2} x^\top Q x + b^\top x + c^\top (z^+ + \mathbf{1} - z^-) \\ \textbf{subject to} \; & \ell_i (1 - z_i^-) \le x_i \le u_i z_i^+ \; \forall i \\ & \cancel{z^- \ge z^+}, \; z^+, z^- \in \{0, 1\}^n \end{aligned} \qquad (\star)$$

A mixed-integer submodular minimization problem! $\Rightarrow$

General case

For simplicity, assume $\ell < \mathbf{0} < u$. Substituting out z_i,

$$\begin{aligned} \underset{x,z^+,z^-\in\mathbb{R}^n}{\textbf{minimize}}\ & \frac{1}{2}x^\top Qx + b^\top x + c^\top(z^+ + \mathbf{1} - z^-) \\ \textbf{subject to}\ & \ell_i(1-z_i^-) \le x_i \le u_i z_i^+\ \forall i \\ & \cancel{z^- \ge z^+},\ z^+, z^- \in \{0,1\}^n \end{aligned} \tag{$\star$}$$

A mixed-integer submodular minimization problem! $\Rightarrow$

$$(\star) \Leftrightarrow \underset{(z^+,z^-)\in\{0,1\}^{2n}}{\textbf{minimize}}\ v(z^+,z^-) + c^\top(z^+ + \mathbf{1} - z^-),$$

where

$$v(z^+,z^-) \overset{\text{def}}{=} \underset{x\in\mathbb{R}^n}{\textbf{minimum}} \left\{ \frac{1}{2}x^\top Qx + b^\top x : \ell_i(1-z_i^-) \le x_i \le u_i z_i^+\ \forall i \right\}$$

is binary submodular and can be evaluated by solving a convex program

Implementation

Fact **All** known algorithms for minimizing binary submodular functions

$$\underset{z \in \{0,1\}^n}{\textbf{minimize}} \; v(z)$$

are required to compute extremal basis at each iteration

Extremal basis Assume at each iteration, the current solution is sorted as $\bar{z}_1 \geq \bar{z}_2 \geq \cdots \geq \bar{z}_n$. The extremal basis[6] (EB) is defined as

$$\left\{ v \left(\mathbf{1}_{[k]} \right) \right\}_{k=0}^{n},$$

where $\mathbf{1}_{[k]} \overset{\text{def}}{=} (\underbrace{1, 1, \ldots, 1}_{k \text{ ones}}, 0, \ldots, 0)$

[6] For delivery purpose, EB defined here is equivalent to but slightly different from the standard one in literature

Implementation

Fact **All** known algorithms for minimizing binary submodular functions

$$\underset{z \in \{0,1\}^n}{\textbf{minimize}} \; v(z)$$

are required to compute extremal basis at each iteration

Extremal basis Assume at each iteration, the current solution is sorted as $\bar{z}_1 \geq \bar{z}_2 \geq \cdots \geq \bar{z}_n$. The extremal basis[6] (EB) is defined as

$$\left\{ v\left(\mathbf{1}_{[k]}\right) \right\}_{k=0}^{n},$$

where $\mathbf{1}_{[k]} \overset{\text{def}}{=} (\underbrace{1, 1, \ldots, 1}_{k \text{ ones}}, 0, \ldots, 0)$

Example If $n = 3$ and $\bar{z}_1 \geq \bar{z}_2 \geq \bar{z}_3$, then one needs to compute

$$v(0,0,0),\ v(1,0,0),\ v(1,1,0),\ v(1,1,1)$$

[6] For delivery purpose, EB defined here is equivalent to but slightly different from the standard one in literature

Fast computation of extremal basis

In the context of sparse/robust MRF (assume $\ell = \mathbf{0}$ for simplicity)

$$v\left(\mathbf{1}_{[k]}\right) = \underset{y \in \mathbb{R}^k}{\textbf{minimum}} \ \left\{ f(y_1, y_2, \ldots, y_k, 0, \ldots, 0) : 0 \le y_i \le u_i \ \forall 1 \le i \le k \right\}$$

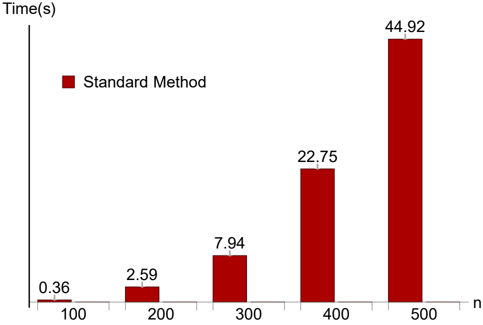
- $f(y) = \frac{1}{2} y^\top Q y + b^\top y \Rightarrow$ convex quadratic program

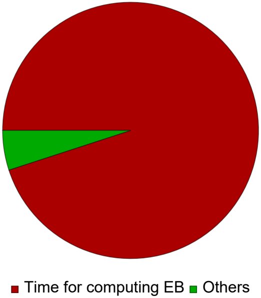
Fast computation of extremal basis

In the context of sparse/robust MRF (assume $\ell = \mathbf{0}$ for simplicity)

$$v\left(\mathbf{1}_{[k]}\right) = \underset{y\in\mathbb{R}^k}{\mathbf{minimum}} \ \{f(y_1, y_2, \ldots, y_k, 0, \ldots, 0) : 0 \le y_i \le u_i \ \forall 1 \le i \le k\}$$

- $f(y) = \frac{1}{2} y^\top Q y + b^\top y \Rightarrow$ solving n QPs per iter! $[\mathcal{O}(n^4)$ operations]





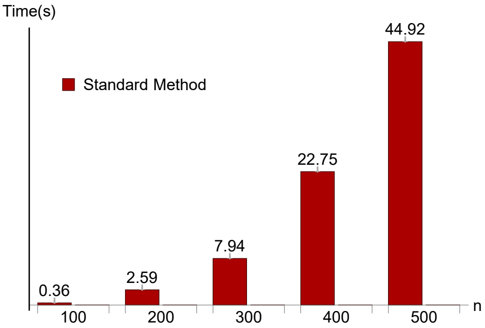
- As $n = 500$, 45 seconds/iter for computing EB $\ge$ 95% total time

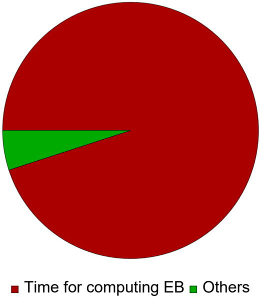
Fast computation of extremal basis

In the context of sparse/robust MRF (assume $\ell = \mathbf{0}$ for simplicity)

$$v\left(\mathbf{1}_{[k]}\right) = \underset{y \in \mathbb{R}^k}{\textbf{minimum}} \ \{f(y_1, y_2, \ldots, y_k, 0, \ldots, 0) : 0 \leq y_i \leq u_i \ \forall 1 \leq i \leq k\}$$

- $f(y) = \frac{1}{2} y^\top Q y + b^\top y \Rightarrow$ solving n QPs per iter! [$\mathcal{O}(n^4)$ operations]





Question: how to efficiently compute $\left\{v\left(\mathbf{1}_{[k]}\right)\right\}_{k=0}^{n}$ in this context?

Fast computation of extremal basis

Idea Assume $\bar{y}^k$ is the optimal solution to k-th subproblem. Consider the parametric optimization problem

$$v_k(y_{k+1}) = \underset{\mathbf{0}\le y\le u_{[k]}}{\textbf{minimum}} \; f(\underbrace{y_1, y_2, \ldots, y_k}_{\text{decision variables}}, \underbrace{y_{k+1}}_{\text{parameter}}, 0, \ldots, 0)$$

$$y^k(y_{k+1}) = \underset{\mathbf{0}\le y\le u_{[k]}}{\arg\min} \; f(y_1, y_2, \ldots, y_k, y_{k+1}, 0, \ldots, 0)$$

Observations

- $y^k(0) = \bar{y}^k$
- $v\left(\mathbf{1}_{[k+1]}\right) = \underset{0\le y_{k+1}\le u_{k+1}}{\textbf{minimum}} \; v_k(y_{k+1})$

Fast computation of extremal basis

Idea Assume $\bar{y}^k$ is the optimal solution to k-th subproblem. Consider the parametric optimization problem

$$v_k(y_{k+1}) = \underset{\mathbf{0} \le y \le u_{[k]}}{\textbf{minimum}} \; f(\underbrace{y_1, y_2, \ldots, y_k}_{\text{decision variables}}, \underbrace{y_{k+1}}_{\text{parameter}}, 0, \ldots, 0)$$

$$y^k(y_{k+1}) = \underset{\mathbf{0} \le y \le u_{[k]}}{\arg\min} \; f(y_1, y_2, \ldots, y_k, y_{k+1}, 0, \ldots, 0)$$

Observations

- $y^k(0) = \bar{y}^k$
- $v\left(\mathbf{1}_{[k+1]}\right) = \underset{0 \le y_{k+1} \le u_{k+1}}{\textbf{minimum}} \; v_k(y_{k+1})$
- $y^k(y'_{k+1}) \le y^k(y''_{k+1})$ if $y'_{k+1} \le y''_{k+1}$. (Isotonicity)

Strategy Increase y_{k+1} from 0 and track $y^k(y_{k+1})$ until find optimal y_{k+1}

Fast computation of extremal basis

Example Consider

$$f(x) = \frac{1}{2}x^\top \begin{bmatrix} 5 & -1 & -3 \\ -1 & 3 & -2 \\ -3 & -2 & 7 \end{bmatrix} x - \sum_{i=1}^{3} x_i, \quad \ell = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad u = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Trajectory of x_1, x_2 and x_3 in terms of driving variables

Fast computation of extremal basis

Example Consider

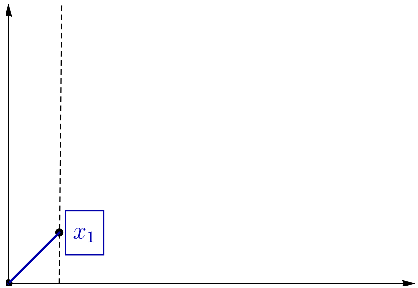
$$f(x) = \frac{1}{2}x^\top \begin{bmatrix} 5 & -1 & -3 \\ -1 & 3 & -2 \\ -3 & -2 & 7 \end{bmatrix} x - \sum_{i=1}^{3} x_i, \quad \ell = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad u = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Trajectory of x_1, x_2 and x_3 in terms of driving variables

- Step 1: to compute

  $$v(1,0,0) = \underset{0 \le x \le 1}{\textbf{minimum}} \ f(x_1, 0, 0),$$

  x_1 is increased from 0



Fast computation of extremal basis

Example Consider

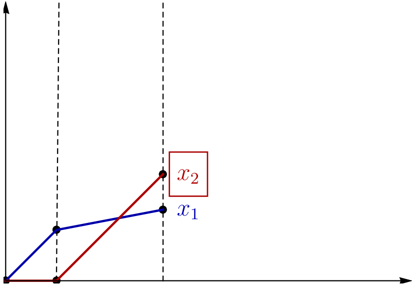
$$f(x) = \frac{1}{2}x^\top \begin{bmatrix} 5 & -1 & -3 \\ -1 & 3 & -2 \\ -3 & -2 & 7 \end{bmatrix} x - \sum_{i=1}^{3} x_i, \quad \ell = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad u = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Trajectory of x_1, x_2 and x_3 in terms of driving variables

- Step 2: to compute

$$v(1,1,0) = \underset{0 \le x \le 1}{\textbf{minimize}} \; f(x_1, x_2, 0),$$

use x_2 to drive the increase of x_1



Fast computation of extremal basis

Example Consider

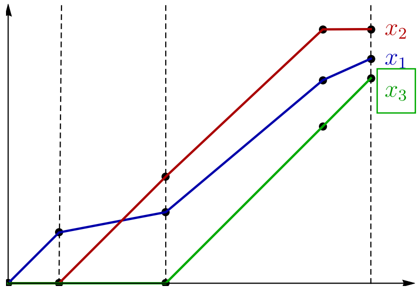
$$f(x) = \frac{1}{2}x^\top \begin{bmatrix} 5 & -1 & -3 \\ -1 & 3 & -2 \\ -3 & -2 & 7 \end{bmatrix} x - \sum_{i=1}^{3} x_i, \quad \ell = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad u = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Trajectory of x_1, x_2 and x_3 in terms of driving variables

- Step 3: to compute

  $$v(1,1,1) = \underset{0 \le x \le 1}{\textbf{minimize}} \; f(x_1, x_2, x_3),$$

  use x_3 to drive the increase of x_1 and x_2



Fast computation of extremal basis

Example Consider

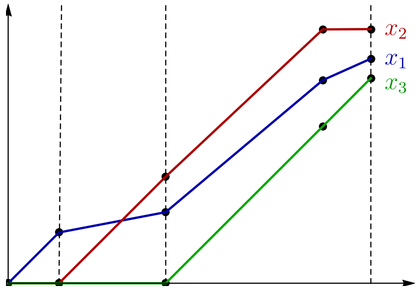
$$f(x) = \frac{1}{2}x^\top \begin{bmatrix} 5 & -1 & -3 \\ -1 & 3 & -2 \\ -3 & -2 & 7 \end{bmatrix} x - \sum_{i=1}^{3} x_i, \quad \ell = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad u = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Trajectory of x_1, x_2 and x_3 in terms of driving variables

- All subproblems

  $v(1,0,0)$, $v(1,1,0)$ and $v(1,1,1)$

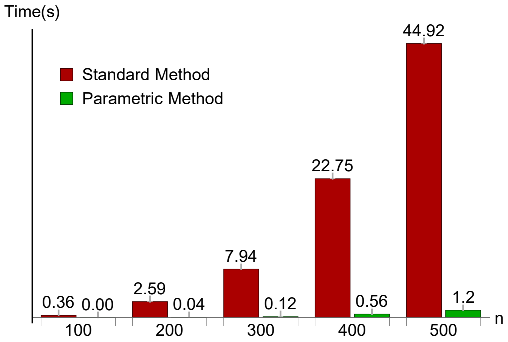
  are solved



Fast computation of extremal basis

Proposition

With fast computation strategy, in each iteration, the sequence $\left\{ v\left(\mathbf{1}_{[k]}\right)\right\}_{i=0}^{n}$ can be computed in $\mathcal{O}(n^3)$.



- 44.92 seconds v.s. 1.2 seconds: $\approx$ 40x faster!

Extension

Results are applicable to many other obj with submodular structures

	Objective $f(x)$	Condition
convex diff	$g(x_i - x_j)$	$g(\bullet)$ convex
conic quadratic	$\sqrt{x^\top Qx}$	$Q_{ij} \leq 0$ & ...
rotated conic quadratic	$\|x\|_2^2/x_0$	$x_0 \geq 0$
Log-Exp	$\log\left(\sum_{i=1}^{n} \exp(x_i)\right)$	–
capped piecewise linear	$\sum_{i=1}^{n} \min\{(a^i)^\top x, b_i\}$	$a^i \geq 0$

- May need additional transformation techniques
- Can appear as substructures in applications, e.g. time-varying regression problems (Bertsimas et al. 2021), mean-risk problems, etc.

Extension

Results are applicable to many other obj with submodular structures

	Objective $f(x)$	Condition
convex diff	$g(x_i - x_j)$	$g(\bullet)$ convex
conic quadratic	$\sqrt{x^\top Qx}$	$Q_{ij} \leq 0$ & ...
rotated conic quadratic	$\|x\|_2^2/x_0$	$x_0 \geq 0$
Log-Exp	$\log\left(\sum_{i=1}^{n} \exp(x_i)\right)$	–
capped piecewise linear	$\sum_{i=1}^{n} \min\{(a^i)^\top x, b_i\}$	$a^i \geq 0$

- May need additional transformation techniques
- Can appear as substructures in applications, e.g. time-varying regression problems (Bertsimas et al. 2021), mean-risk problems, etc.
$\Rightarrow$ How to exploit submodularity?

Extension

Results are applicable to many other obj with submodular structures

	Objective $f(x)$	Condition
convex diff	$g(x_i - x_j)$	$g(\bullet)$ convex
conic quadratic	$\sqrt{x^\top Qx}$	$Q_{ij} \leq 0$ & ...
rotated conic quadratic	$\|x\|_2^2/x_0$	$x_0 \geq 0$
Log-Exp	$\log\left(\sum_{i=1}^{n} \exp(x_i)\right)$	–
capped piecewise linear	$\sum_{i=1}^{n} \min\{(a^i)^\top x, b_i\}$	$a^i \geq 0$

- May need additional transformation techniques
- Can appear as substructures in applications, e.g. time-varying regression problems (Bertsimas et al. 2021), mean-risk problems, etc. $\Rightarrow$ How to exploit submodularity? Convexification

Recap

Summary

- Sparse/robust MRF inference problems are **polynomially solvable**!
- Fast computation of extremal basis
- The computational approach is efficient in practice

Recap

Summary

- Sparse/robust MRF inference problems are **polynomially solvable**!
- Fast computation of extremal basis
- The computational approach is efficient in practice

Thanks for your listening!

Reference I

Bertsimas, D., Digalakis Jr, V., Li, M. L., and Lami, O. S. (2021). Slowly varying regression under sparsity. arXiv preprint arXiv:2102.10773.

Elmachtoub, A. N. and Grigas, P. (2022). Smart "predict, then optimize". Management Science, 68(1):9–26.

Law, J., Quick, M., and Chan, P. (2014). Bayesian spatio-temporal modeling for analysing local patterns of crime over time at the small-area level. Journal of quantitative criminology, 30:57–78.

Morris, M., Wheeler-Martin, K., Simpson, D., Mooney, S. J., Gelman, A., and DiMaggio, C. (2019). Bayesian hierarchical spatial models: Implementing the besag york mollié model in stan. Spatial and spatio-temporal epidemiology, 31:100301.

Schrunner, S., Bluder, O., Zernig, A., Kaestner, A., and Kern, R. (2017). Markov random fields for pattern extraction in analog wafer test data. In 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), pages 1–6. IEEE.

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