Approximation guarantees
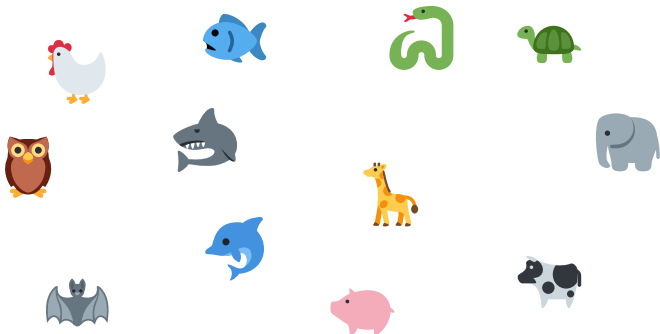# of local search algorithms
via localizability of set functions

Kaito Fujii

(National Institute of Informatics)

A function that assigns a value to every subset

A function that assigns a value to every subset

$$f(\{\; \text{🐬}, \text{🐍}, \text{🐘}\;\}) = 50$$

A function that assigns a value to every subset

$$f(\{\text{🐄}, \text{🦒}\}) = 20$$

Finding a subset maximizing the objective value
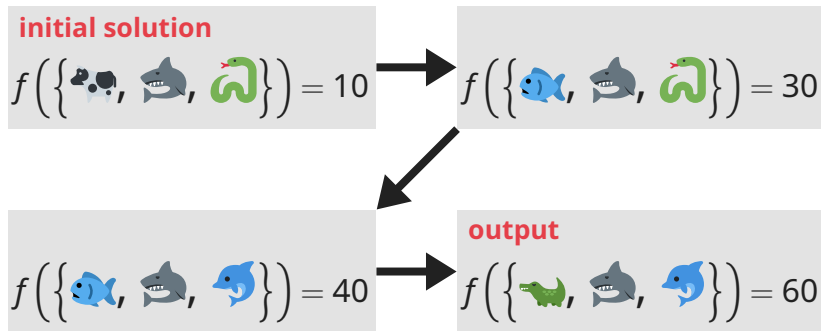
Maximize $f(X)$

subject to $X \in \mathcal{I}$

$f : 2^N \rightarrow \mathbb{R}$ set function

N is the ground set

Algorithmic framework for set function max.

initial solution

$f\left(\{🐩, 🦈, 🐍\}\right) = 10$

$f\left(\{🐟, 🦈, 🐍\}\right) = 30$

$f\left(\{🐟, 🦈, 🐬\}\right) = 40$

output

$f\left(\{🐊, 🦈, 🐬\}\right) = 60$

Q **When do local search algorithms work well?**

We propose a property called *localizability*

We propose a property called *localizability*

1 If the objective function satisfies localizability, local search is guaranteed to work well

We propose a property called *localizability*

1 If the objective function satisfies localizability, local search is guaranteed to work well

2 The objective function of sparse optimization satisfies localizability

Our contribution

We propose a property called *localizability*

1 If the objective function satisfies localizability, local search is guaranteed to work well

2 The objective function of sparse optimization satisfies localizability

3 Local search for sparse optimization can be accelerated

Finding a subset of size at most s

Maximize $f(X)$

subject to $|X| \leq s$

Finding a subset of size at most s

Maximize $f(X)$

subject to $|X| \leq s$

Assume f is monotone

i.e. $A \subseteq B \Rightarrow f(A) \leq f(B)$

Warmup: Cardinality constraints

Finding a subset of size at most s

Maximize $f(X)$ — Assume f is monotone

i.e. $A \subseteq B \Rightarrow f(A) \leq f(B)$

subject to $|X| \leq s$

An algorithm is α-approximation ($\alpha \in [0, 1]$)

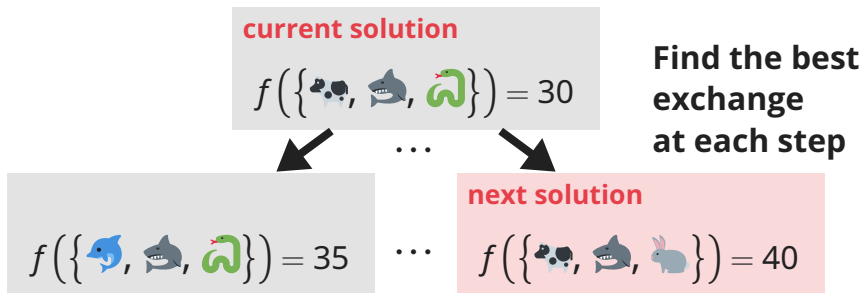$\overset{\triangle}{\Leftrightarrow}$ $f(X) \geq \alpha f(X^*)$,

where X is the output and X^* is an optimal

Let X be any maximal feasible solution
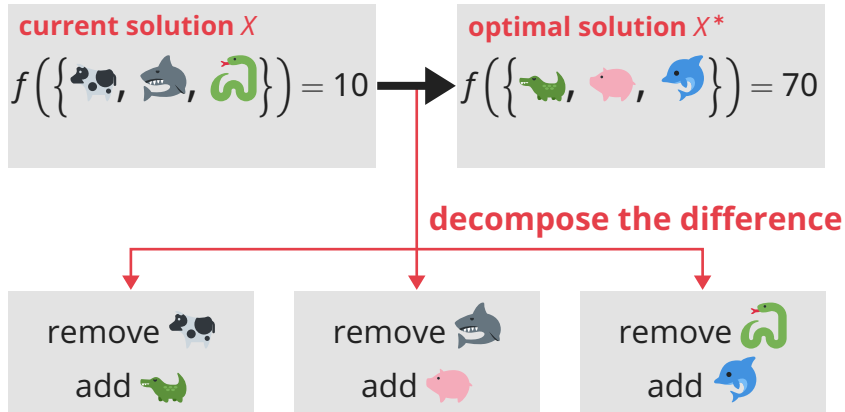
For $i = 1, \cdots, T$:

    For each $a \in N \setminus X, \ b \in X$, compute $f(X \setminus \{b\} \cup \{a\})$

    Update X with the best exchange

current solution

$f\left(\{🐘, 🦈, 🐍\}\right) = 30$

Find the best exchange at each step

$f\left(\{🐬, 🦈, 🐍\}\right) = 35$ $\cdots$ **next solution** $f\left(\{🐘, 🦈, 🐰\}\right) = 40$

current solution X

$$f\left(\left\{\text{🐄}, \text{🦈}, \text{🐍}\right\}\right) = 10$$

optimal solution X^*

$$f\left(\left\{\text{🐊}, \text{🐷}, \text{🐬}\right\}\right) = 70$$

decompose the difference

remove 🐄
add 🐊

remove 🦈
add 🐷

remove 🐍
add 🐬

f is (α, β)-localizable
$\overset{\triangle}{\Longleftrightarrow}$ **sum of local improvement** $\geq \alpha f(X^*) - \beta f(X)$

Theorem

If the objective function is (α, β)-localizable,

our local search algo. is $\dfrac{\alpha}{\beta}\left(1 - \exp\left(-\dfrac{\beta T}{s}\right)\right)$-approx.

T: #iterations, s: maximum solution size

Theorem

If the objective function is (α, β)-localizable,

our local search algo. is $\dfrac{\alpha}{\beta}\left(1 - \exp\left(-\dfrac{\beta T}{s}\right)\right)$-approx.

T: #iterations, s: maximum solution size

Guarantees for well-known classes

	(α, β)	approx. guarantee
linear	$(1, 1)$	$(1 - \exp(-T/s))$
submodular	$(1, 2)$	$\frac{1}{2}(1 - \exp(-T/s))$

Maximize $f(X)$

subject to $X \in \mathcal{I}$

$\mathcal{I} \subseteq 2^N$ represents a family of feasible subsets

Various classes of constraints

p-exchange systems

$\not\subseteq$
$\not\supseteq$

p-matroid intersection

$\cup$I

matroid

$\cup$I

cardinality
$\mathcal{I} = \{X \subseteq N \mid |X| \leq s\}$

N finite set, $\mathcal{I} \subseteq 2^N$ s.t. $\varnothing \in \mathcal{I}$ and $A \subseteq B \in \mathcal{I} \Rightarrow A \in \mathcal{I}$
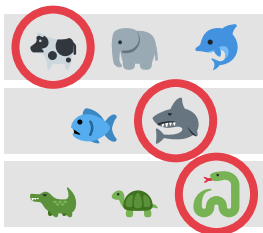
$\mathcal{M} = (N, \mathcal{I})$ is a **matroid**
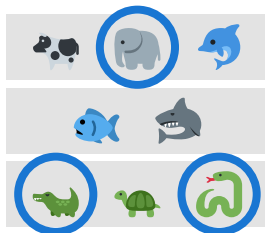
⬍ **def**

$\forall A, B \in \mathcal{I}, \ |A| < |B| \Rightarrow \exists i \in B \setminus A, \ A \cup \{i\} \in \mathcal{I}$

e.g.) partition matroid

independent   **dependent**

N finite set, $\mathcal{I} \subseteq 2^N$ s.t. $\varnothing \in \mathcal{I}$ and $A \subseteq B \in \mathcal{I} \Rightarrow A \in \mathcal{I}$

$\mathcal{M} = (N, \mathcal{I})$ is a **matroid**

⬍ **def**

$\forall A, B \in \mathcal{I}, \ |A| < |B| \Rightarrow \exists i \in B \setminus A, \ A \cup \{i\} \in \mathcal{I}$

$(N, \mathcal{I})$ is a *p***-matroid intersection**

⬍ **def**

$\exists (N, \mathcal{I}_1), \cdots, (N, \mathcal{I}_p)$ matroids s.t. $\mathcal{I} = \bigcap_{i=1}^{p} \mathcal{I}_p$

N finite set, $\mathcal{I} \subseteq 2^N$ s.t. $\varnothing \in \mathcal{I}$ and $A \subseteq B \in \mathcal{I} \Rightarrow A \in \mathcal{I}$
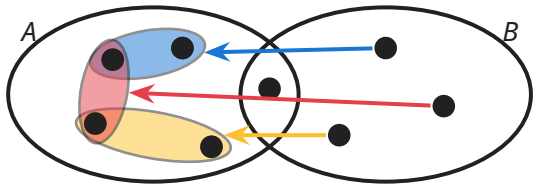
$(N, \mathcal{I})$ is a *p-exchange system*

⬍ **def**

$\forall A, B \in \mathcal{I}$

$\exists \phi : B \setminus A \to 2^{A \setminus B}$ s.t.

● For all $B' \subseteq B \setminus A$, $(A \setminus (\bigcup_{v \in B'} \phi(v))) \cup B' \in \mathcal{I}$

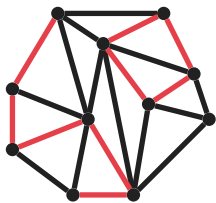● $|\phi(v)| \leq p \ (\forall v \in B \setminus A)$

● Each $v' \in A \setminus B$ appears at most p sets of $(\phi(v))_{v \in B \setminus A}$
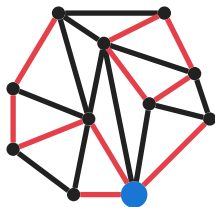
N finite set, $\mathcal{I} \subseteq 2^N$ s.t. $\emptyset \in \mathcal{I}$ and $A \subseteq B \in \mathcal{I} \Rightarrow A \in \mathcal{I}$

$(N, \mathcal{I})$ is a *p*-**exchange system**

e.g.) *b*-matching
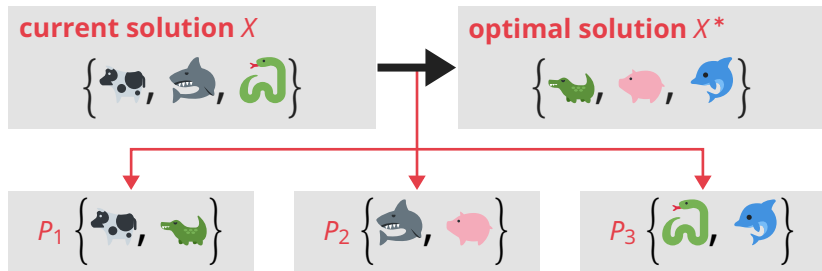

2-**matching**   **not** 2-**matching**

$\mathcal{I} = \{X \subseteq E \mid \forall v \in V, \ \deg_v(X) \leq b\}$ is a 2-exchange sys.

current solution X

$$\left\{ \text{🐄}, \text{🦈}, \text{🐍} \right\}$$

optimal solution X^*

$$\left\{ \text{🐊}, \text{🐷}, \text{🐬} \right\}$$

$P_1 \left\{ \text{🐄}, \text{🐊} \right\}$ $P_2 \left\{ \text{🦈}, \text{🐷} \right\}$ $P_3 \left\{ \text{🐍}, \text{🐬} \right\}$

$\mathcal{P}$ a multiset of 2^N s.t. $\begin{cases} \text{each of } X^* \setminus X \text{ appears } k \text{ times} \\ \text{each of } X \setminus X^* \text{ appears } \ell \text{ times} \end{cases}$

f is $(\alpha, \beta_1, \beta_2)$-localizable
$$\overset{\triangle}{\Longleftrightarrow} \sum_{P \in \mathcal{P}} \left\{ f(X \triangle P) - f(X) \right\} \geq \alpha k \, f(X^*) - (\beta_1 \ell + \beta_2 k) f(X)$$

Theorem

If the objective function is $(\alpha, \beta_1, \beta_2)$-localizable

Matroids the local search algorithm is

$$\frac{\alpha}{\beta_1 + \beta_2}\left(1 - \exp\left(-\frac{(\beta_1 + \beta_2)T}{s}\right)\right)\text{-approximation}$$

p-MI/p-ES the local search algo. with paramter

$$q \in \mathbb{Z}_{>0} \text{ is } \frac{\alpha\left(1 - \exp\left(\frac{(\beta_1(p-1+1/q)+\beta_2)T}{s}\right)\right)}{\beta_1(p-1+1/q)+\beta_2}\text{-approx.}$$

q is a parameter of the algorithms you can choose
(you must check $n^{O(q)}$ solutions at each step)

Finding a sparse solution for continuous opt.

$$\text{Maximize}_{\mathbf{w} \in \mathbb{R}^n} \quad u(\mathbf{w})$$

$$\text{subject to} \quad \text{supp}(\mathbf{w}) \in \mathcal{I}$$

Notation

- $u \colon \mathbb{R}^n \to \mathbb{R}$ continuous function with $u(\mathbf{0}) \geq 0$
- $N \triangleq \{1, \cdots, n\}$
- $\text{supp}(\mathbf{w}) \triangleq \{i \in N \mid \mathbf{w}_i \neq 0\}$ indices of non-zeros

Finding a sparse solution for continuous opt.

Maximize$_{\mathbf{w} \in \mathbb{R}^n}$ $u(\mathbf{w})$

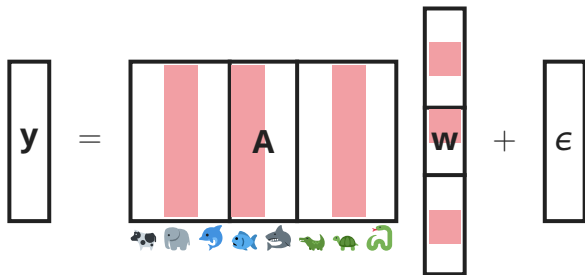subject to $\text{supp}(\mathbf{w}) \in \mathcal{I}$ ◄ Structured constraints

Notation

- $u \colon \mathbb{R}^n \to \mathbb{R}$ continuous function with $u(\mathbf{0}) \geq 0$
- $N \triangleq \{1, \cdots, n\}$
- $\text{supp}(\mathbf{w}) \triangleq \{i \in N \mid \mathbf{w}_i \neq 0\}$ indices of non-zeros

Selecting one feature from each category
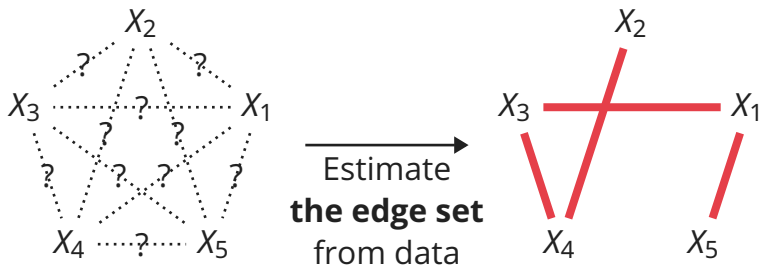

$$\text{Maximize} \quad u_{R^2}(\mathbf{w}) \triangleq 1 - \frac{\|\mathbf{y} - \mathbf{A}\mathbf{w}\|_2^2}{\|\mathbf{y}\|_2^2}$$

$$\text{subject to} \quad \text{supp}(\mathbf{w}) \in \mathcal{I}$$

Estimate supp($\mathbf{w}$) **of a sparse Ising model**

$$p(\mathbf{x}|\mathbf{w}) \propto \exp\left(\sum_{(u,v)\in E} w_{uv}x_u x_v + \sum_{u\in V} w_u x_u\right)$$

from data $\{\mathbf{x}^1, \cdots, \mathbf{x}^m\} \sim p(\mathbf{x}|\mathbf{w})$
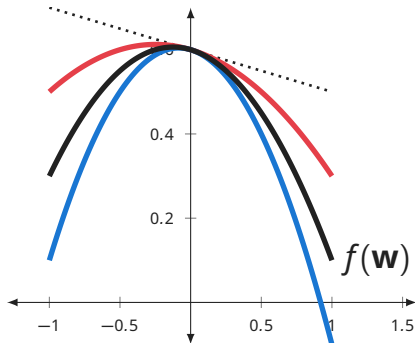

A degree constraint is a special case of a 2-ES constr.

u is restricted strongly concave

$$\Leftrightarrow \forall \mathbf{x}, \mathbf{y} \in \Omega, \ u(\mathbf{y}) - u(\mathbf{x}) - \langle \nabla u(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \leq -\frac{m_\Omega}{2} \|\mathbf{y} - \mathbf{x}\|_2^2$$

u is restricted smooth

$$\Leftrightarrow \forall \mathbf{x}, \mathbf{y} \in \Omega, \ u(\mathbf{y}) - u(\mathbf{x}) - \langle \nabla u(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \geq -\frac{M_\Omega}{2} \|\mathbf{y} - \mathbf{x}\|_2^2$$

Restricted strong concavity constant m_s on $\Omega_s = \{\mathbf{x}, \mathbf{y} \mid \|\mathbf{x}\|_0 \leq s, \ \|\mathbf{y}\|_0 \leq s, \|\mathbf{x} - \mathbf{y}\|_0 \leq s\}$

Restricted smoothness constant $M_{s,t}$ on $\Omega_{s,t} = \{\mathbf{x}, \mathbf{y} \mid \|\mathbf{x}\|_0 \leq s, \ \|\mathbf{y}\|_0 \leq s, \|\mathbf{x} - \mathbf{y}\|_0 \leq t\}$

We can reduce sparse opt. to set function opt.

$$\text{Maximize}_{\mathbf{w}\in\mathbb{R}^n} \quad u(\mathbf{w})$$
$$\text{subject to} \quad \text{supp}(\mathbf{w}) \in \mathcal{I}$$

⬇ Define $f_u(X) \stackrel{\triangle}{=} \max\limits_{\mathbf{w}:\,\text{supp}(\mathbf{w})\subseteq X} u(\mathbf{w})$

$$\text{Maximize}_{X\subseteq N} \quad f_u(X)$$
$$\text{subject to} \quad X \in \mathcal{I}$$

f_u is $\left(\frac{m_{2s}}{M_{s,t}}, \frac{M_{s,t}}{m_{2s}}, 0\right)$-localizable with size s and exchange size t

Approxima. guarantees for sparse opt.

Constraint	Local search	Greedy
Cardinality	$\frac{m_{2s}^2}{M_{s,2}^2}(1-\epsilon_1(T))$	$1-\exp\left(-\frac{m_{2s}}{M_{s,1}}\right)$ †
Matroids	$\frac{m_{2s}^2}{M_{s,2}^2}(1-\epsilon_1(T))$	$\frac{1}{(1+\frac{M_{s,1}}{m_s})^2}$ ‡
p-MI/p-ES	$\frac{1}{p-1+1/q}\frac{m_{2s}^2}{M_{s,2}^2}(1-\epsilon_2(T))$	N/A

$\epsilon_1(T)$ and $\epsilon_2(T)$ are terms converging to 0 as $T \to \infty$

† [Elenberg–Khanna–Dimakis–Negahban'18]

‡ [Chen–Feldman–Karbasi'18]

It takes much time to decide $\underset{a \in N \setminus X, b \in X}{\operatorname{argmax}} f_u(X \setminus \{a\} \cup \{b\})$

➡ **Approximately evaluate the value of f_u**

It takes much time to decide $\displaystyle\operatorname*{argmax}_{a \in N \setminus X, b \in X} f_u(X \setminus \{a\} \cup \{b\})$

➡ **Approximately evaluate the value of f_u**

$$\mathbf{w}^{(X)} \in \operatorname*{argmax}_{\mathbf{w}\,:\,\mathrm{supp}(\mathbf{w}) \subseteq X} u(\mathbf{w})$$

semi-oblivious

$\displaystyle\operatorname*{argmax}_{a \in N \setminus X} f_u(X \setminus \{b\} \cup \{a\})$, where $b \in \operatorname*{argmin}_{b \in X}(\mathbf{w}^{(X)})_b^2$

Quickly decides the element to be removed

It takes much time to decide $\underset{a \in N \setminus X, b \in X}{\mathrm{argmax}} \, f_u(X \setminus \{a\} \cup \{b\})$

➡ **Approximately evaluate the value of f_u**

$$\mathbf{w}^{(X)} \in \underset{\mathbf{w}:\, \mathrm{supp}(\mathbf{w}) \subseteq X}{\mathrm{argmax}} \, u(\mathbf{w})$$

semi-oblivious

$\underset{a \in N \setminus X}{\mathrm{argmax}} f_u(X \setminus \{b\} \cup \{a\})$, where $b \in \underset{b \in X}{\mathrm{argmin}} (\, \mathbf{w}^{(X)} \,)_b^2$

Quickly decides the element to be removed

non-oblivious

$$\underset{a \in N \setminus X, b \in X}{\mathrm{argmax}} \left\{ \frac{1}{2M_{s,2}} \left(\nabla u(\mathbf{w}^{(X)}) \right)_a^2 - \frac{M_{s,2}}{2} \left(\mathbf{w}^{(X)} \right)_b^2 \right\}$$

We propose a property called *localizability*

1 If the objective function satisfies localizability, local search is guaranteed to work well

2 The objective function of sparse optimization satisfies localizability

3 Local search for sparse optimization can be accelerated