Information-theoretic coordinate subset and partition selection of multivariate Markov chains via submodular optimization

Zheyuan Lai

Joint work with Professor Michael Choi



Department of Statistics
and Data Science
Faculty of Science

San Francisco, CA
Nov 1, 2026

Introduction

- Based on Zheyuan Lai and Michael C. H. Choi, "Information-theoretic coordinate subset and partition selection of multivariate Markov chains via submodular optimization", *Journal of Combinatorial Optimization*, 2026.
- A discrete multivariate Markov chain P on $\mathcal{X} = \mathcal{X}^{(1)} \times \cdots \times \mathcal{X}^{(d)}$ with stationary distribution π.

  $P(x, y)$: probability of moving from x to y; $\quad \pi P = \pi$.

  Example: d spins $x^i \in \{-1, +1\}$ of a magnet (Curie–Weiss model).
- Dimension reduction: Which m of the d coordinates should we keep?
- We use submodularity to design greedy algorithms with guarantees.
- Framework: Michael C. H. Choi, Youjia Wang and Geoffrey Wolfer, "Geometry and factorization of multivariate Markov chains with applications to MCMC acceleration and approximate inference", To appear in *SIAM/ASA Journal on Uncertainty Quantification*, 2026.

Projection and information theory of Markov chains

Definition 1 (Keep-S-in (resp. leave-S-out) transition matrix)

For $S \subseteq [\![d]\!]$, we project the original Markov chain to S via

$$P^{(S)}(x^S, y^S) := \frac{\sum_{x^{-S}, y^{-S}} \pi(x)\, P(x,y)}{\sum_{x^{-S}} \pi(x)}.$$

Definition 2 (Entropy rate of a Markov chain)

$$H(P) := -\sum_{x,y} \pi(x)\, P(x,y) \ln P(x,y)$$

Definition 3 (KL divergence ("Distance") between Markov chains)

$$D(M \,\|\, L) := \sum_x \pi(x) \sum_y M(x,y) \ln \frac{M(x,y)}{L(x,y)}$$

Definition 4 (Distance to independence)

$$\mathbb{I}(P) := \min_{L_i} D\left(P \,\|\, \otimes_{i=1}^d L_i\right) = D\left(P \,\|\, \otimes_{i=1}^d P^{(i)}\right)$$

Submodularity: diminishing returns

Definition 5 (Submodular function)

$f : 2^U \to \mathbb{R}$ is submodular if for all $S \subseteq T \subseteq U$ and $e \notin T$,

$$f(S \cup \{e\}) - f(S) \;\geq\; f(T \cup \{e\}) - f(T).$$

It is monotonically nondecreasing if $f(S) \leq f(T)$ whenever $S \subseteq T$.

Submodular maximization subject to cardinality constraint:

$$\max_{|S| \leq m} f(S).$$

Theorem 6 (Nemhauser, Wolsey and Fisher '78)

If f is monotonically nondecreasing and submodular with $f(\emptyset) = 0$, the greedy algorithm returns S with $f(S) \geq (1 - e^{-1}) f(\mathrm{OPT})$.

Submodular functions in Markov chain theory

Theorem 7 (Choi, Wang and Wolfer '26)

Let $\pi = \otimes_{i=1}^{d} \pi^{(i)}$ be a product stationary distribution. Then

1. $S \mapsto H(P^{(S)})$ *is submodular but non-monotone;*
2. $S \mapsto D(P \| P^{(S)} \otimes P^{(-S)})$ *is submodular but non-monotone;*
3. $S \mapsto \mathbb{I}(P^{(S)})$ *is monotonically nondecreasing and supermodular.*

Open problem from Choi, Wang and Wolfer '26:

Can we design combinatorial optimization algorithms to find some "optimal" Markov chains subject to cardinality constraints?

This work: more submodular functions, and algorithms with guarantees.

New supermodular functions and k-submodularity

Theorem 8 (L. and Choi '26)

Let π be of product form and every row of Π be π. Then

1. $S \mapsto \mathbb{I}(P^{(-S)})$ *is nonincreasing and supermodular;*
2. $S \mapsto D(P^{(S)} \,\|\, \Pi^{(S)})$ *is nondecreasing and supermodular.*

Definition 9 (k-submodularity)

Each element of U goes into one of k groups or none: $S = (S_1, \ldots, S_k)$.
$\Delta_{e,j} f(S)$: the gain from adding e to group j.
f is k-submodular iff it is (Ward and Živný '16)

- orthant submodular: $\Delta_{e,j} f(S) \geq \Delta_{e,j} f(T)$ whenever $S_i \subseteq T_i$ for all i;
- pairwise monotone: $\Delta_{e,i} f(S) + \Delta_{e,j} f(S) \geq 0$ for $i \neq j$.

Overview of results

Objective		Structure	Algorithm	Guarantee
$\max H(P^{(S)})$	§3	submodular	distorted greedy	$(1-\frac{1}{e})\,g-c$
$\max H(P^{(S)})$	§3	submodular	local search	$(\frac{1}{3}-\frac{\epsilon}{d})$-approx.
$\max D(P \,\|\|\, P^{(S)} \otimes P^{(-S)})$	§4	submodular, symmetric	distorted greedy	$(1-\frac{1}{e})\,g-c$
$\max D(P \,\|\|\, P^{(S)} \otimes P^{(-S)})$	§4	submodular, symmetric	local search	$(\frac{1}{2}-\frac{\epsilon}{d})$-approx.
$\min \mathbb{I}(P^{(S)})$	§5	supermodular, increasing	distorted greedy	$(1-\frac{1}{e})\,g-c$
$\min \mathbb{I}(P^{(-S)})$	§5.1	supermodular, decreasing	greedy	$(1-\frac{1}{e})$-approx.
$\max D(P^{(S)} \,\|\|\, \Pi^{(S)})$	§6	increasing	batch greedy	ratio bound
$\min D(P^{(-S)} \,\|\|\, \Pi^{(-S)})$	§6.1	supermodular, decreasing	greedy	$(1-\frac{1}{e})$-approx.
optimization on partition	§3–6	orthant submodular	generalized distorted greedy	$(1-\frac{1}{e})\,g-c$

What if f is not monotone?

Transform $f = g - c$ with g monotone submodular and $c \geq 0$ modular

Proposition 10 (Korte and Vygen '11)

$g(S) := f(S) + \sum_{e \in S}(f(U \setminus \{e\}) - f(U))$

Algorithm 1 Distorted greedy (Harshaw, Feldman, Ward and Karbasi '19)

Require: monotone submodular g, modular $c \geq 0$, cardinality constr. m

1: $S_0 \leftarrow \emptyset$
2: **for** $i = 0, \ldots, m-1$ **do**
3: $\quad e_i \leftarrow \arg\max_{e \in U} \left\{ \left(1 - \frac{1}{m}\right)^{m-i-1} \left[g(S_i \cup \{e\}) - g(S_i)\right] - c(\{e\}) \right\}$
4: $\quad S_{i+1} \leftarrow S_i \cup \{e_i\}$ if this maximum is positive, else $S_{i+1} \leftarrow S_i$
5: **end for**
6: **return** S_m

Theorem 11 (Harshaw, Feldman, Ward and Karbasi '19)

$g(S_m) - c(S_m) \geq (1 - e^{-1})\, g(\mathrm{OPT}) - c(\mathrm{OPT})$

From orthant submodularity to k-submodularity

Theorem 12 (L. and Choi '26)

Let f be orthant submodular and $V = (V_1, \ldots, V_k)$. On $\{S : S_j \subseteq V_j \ \forall j\}$,

$$g(S) := f(S) + \sum_{j=1}^{k} \sum_{e \in S_j} \left[f(V_1, \ldots, V_j \setminus \{e\}, \ldots, V_k) - f(V) \right]$$

is monotone and k-submodular. Hence $f = g - c$ with c modular.

Generalized distorted greedy algorithm

Algorithm 2 Generalized distorted greedy (L. and Choi '26)

Require: monotone k-submodular g, modular $c \geq 0$, budget m, groups V

1: $S_0 \leftarrow (\emptyset, \ldots, \emptyset)$
2: **for** $i = 0, \ldots, m-1$ **do**
3: $\quad (j^*, e^*) \leftarrow \arg\max_{j,\, e \in V_j \setminus S_{i,j}} \left\{ \left(1 - \frac{1}{m}\right)^{m-i-1} \Delta_{e,j}\, g(S_i) - c(\{e\}) \right\}$
4: $\quad$ add e^* to group j^* if this maximum is positive
5: **end for**
6: **return** $S_m = (S_{m,1}, \ldots, S_{m,k})$

Theorem 13 (L. and Choi '26)

$g(S_m) - c(S_m) \geq (1 - e^{-1})\, g(\mathrm{OPT}) - c(\mathrm{OPT})$ *over* $S_j \subseteq V_j$, $|\cup_j S_j| \leq m$.

Application to MCMC: splitting off one coordinate

- Setup: Curie–Weiss model with $d = 8$ spins.
- Aim: design accelerated MCMC algorithm via state space projection.
- By submodular optimization, select

$$i^\star = \arg\min_i D(P^{(-i)} \,\|\, \Pi^{(-i)}) = 4.$$

- (Empirically) Run $P^{(-4)} \otimes P^{(4)}$ in place of P:

Markov chain	TV to π after 10 steps
P	0.217
$P^{(-4)} \otimes P^{(4)}$	**0.187**
worst split ($i = 1$ or 8)	0.200

Takeaways

Contributions

1. "Dimension reduction" of multivariate Markov chains is submodular optimization of information-theoretic criteria.
2. Optimization theory for orthant submodular functions: a transform and a generalized distorted greedy algorithm with the same $(1 - e^{-1})$-type lower bound.

Future directions

Extend the current framework and algorithms to

1. continuous state space
2. larger $d \Longrightarrow$ language models?

Thank you! Question(s)?

Lai and Choi, *J. Comb. Optim.* 51, 40 (2026)

References

M. C. H. Choi, Y. Wang and G. Wolfer. Geometry and factorization of multivariate Markov chains with applications to MCMC acceleration and approximate inference. To appear in *SIAM/ASA Journal on Uncertainty Quantification*, 2026.

C. Harshaw, M. Feldman, J. Ward and A. Karbasi. Submodular maximization beyond non-negativity: guarantees, fast algorithms, and applications. *ICML*, 2019.

B. Korte and J. Vygen. *Combinatorial Optimization*. Springer, 2011.

Z. Lai and M. C. H. Choi. **Information-theoretic coordinate subset and partition selection of multivariate Markov chains via submodular optimization.** *J. Comb. Optim.* 51:40, 2026.

G. L. Nemhauser, L. A. Wolsey and M. L. Fisher. An analysis of approximations for maximizing submodular set functions I. *Math. Programming* 14:265–294, 1978.

J. Ward and S. Živný. Maximizing k-submodular functions and beyond. *ACM Trans. Algorithms* 12(4):47, 2016.