Bregman Proximal Methods for Quantum Information Theoretic Problems

K. He[1] J. Saunderson[1] H. Fawzi[2]

[1]Department of Electrical and Computer System Engineering
Monash University

[2]Department of Applied Mathematics and Theoretical Physics
University of Cambridge

25th International Symposium on Mathematical Programming
July 2024

Quantum Channel Capacities

Quantum channel capacity (Bennet et al., 1997):

$$\max_{X \in \mathbb{H}^n} \quad I(X) \quad \text{subj. to} \quad \text{tr}[X] = 1, \; X \succeq 0,$$

where

$$I(X) := S(X) + S(\mathcal{N}(X)) - S(\mathcal{M}(X)) \qquad \text{(Quantum mutual inf.)}$$
$$S(X) := -\text{tr}[X \log(X)] \qquad \text{(Quantum entropy)}$$

and $\mathcal{N}, \mathcal{M} : \mathbb{H}^n \to \mathbb{H}^n$ are (related) linear functions.

Quantum Channel Capacities

Quantum channel capacity (Bennet et al., 1997):

$$\max_{X \in \mathbb{H}^n} \quad I(X) \quad \text{subj. to} \quad \text{tr}[X] = 1, \; X \succeq 0,$$

where

$$I(X) := S(X) + S(\mathcal{N}(X)) - S(\mathcal{M}(X)) \qquad \text{(Quantum mutual inf.)}$$

$$S(X) := -\text{tr}[X \log(X)] \qquad \text{(Quantum entropy)}$$

and $\mathcal{N}, \mathcal{M} : \mathbb{H}^n \to \mathbb{H}^n$ are (related) linear functions.

How can we efficiently compute this quantity?

- SDP approx. (Fawzi et al., 2019)? But no practical SDP solver for large-scale problems of this type
- Projected gradient-descent-type algorithms don't work well

Table of Contents

Table of Contents

Preliminaries

For positive definite matrix $X = \sum_{i=1}^{n} \lambda_i v_i v_i^\top$, define **matrix logarithm** as

$$\log(X) = \sum_{i=1}^{n} \log(\lambda_i) v_i v_i^\top$$

Preliminaries

For positive definite matrix $X = \sum_{i=1}^{n} \lambda_i v_i v_i^\top$, define **matrix logarithm** as

$$\log(X) = \sum_{i=1}^{n} \log(\lambda_i) v_i v_i^\top$$

Entropy:

- Classical: $H(x) := -\sum_{i=1}^{n} x_i \log(x_i)$
- Quantum: $S(X) := -\operatorname{tr}[X \log(X)] = H(\lambda)$

Preliminaries

For positive definite matrix $X = \sum_{i=1}^{n} \lambda_i v_i v_i^\top$, define **matrix logarithm** as

$$\log(X) = \sum_{i=1}^{n} \log(\lambda_i) v_i v_i^\top$$

Entropy:

- Classical: $H(x) \coloneqq -\sum_{i=1}^{n} x_i \log(x_i)$
- Quantum: $S(X) \coloneqq -\operatorname{tr}[X \log(X)] = H(\lambda)$

Relative entropy:

- Classical: $H(x \,\|\, y) \coloneqq \sum_{i=1}^{n} x_i \log(x_i / y_i)$
- Quantum: $S(X \,\|\, Y) \coloneqq \operatorname{tr}[X(\log(X) - \log(Y))]$

Quantum relative entropy is jointly convex in X and Y (nontrivial!)

Mirror descent

Consider constrained convex optimization problem

$$\min_{x \in \mathcal{X}} \quad f(x).$$

Mirror descent

Consider constrained convex optimization problem

$$\min_{x \in \mathcal{X}} \quad f(x).$$

Projected gradient descent can be represented as

$$x^{k+1} = \arg\min_{x \in \mathcal{X}} \langle \nabla f(x^k), x \rangle + \frac{1}{2t_k} \|x - x^k\|_2^2$$

Mirror descent

Consider constrained convex optimization problem

$$\min_{x \in \mathcal{X}} \quad f(x).$$

Projected gradient descent can be represented as

$$x^{k+1} = \arg\min_{x \in \mathcal{X}} \langle \nabla f(x^k), x \rangle + \frac{1}{2t_k} \|x - x^k\|_2^2$$

Mirror descent replaces Euclidean norm with Bregman divergence

$$x^{k+1} = \arg\min_{x \in \mathcal{X}} \langle \nabla f(x^k), x \rangle + \frac{1}{t_k} D_\varphi(x \| x^k)$$

where

$$D_\varphi(x \| y) := \varphi(x) - (\varphi(y) + \langle \nabla \varphi(y), x - y \rangle).$$

Mirror descent

Mirror descent with $\mathcal{X}$ probability simplex, $\varphi(x) = -H(x)$ gives

$$x_i^{k+1} = \frac{x_i^k \exp(-t_k \partial_i f(x))}{\sum_{j=1}^n x_j^k \exp(-t_k \partial_j f(x))}, \quad \forall i = 1, \ldots, n.$$

Mirror descent

Mirror descent with $\mathcal{X}$ probability simplex, $\varphi(x) = -H(x)$ gives

$$x_i^{k+1} = \frac{x_i^k \exp(-t_k \partial_i f(x))}{\sum_{j=1}^n x_j^k \exp(-t_k \partial_j f(x))}, \quad \forall i = 1, \ldots, n.$$

Mirror descent with $\mathcal{X}$ unit trace PSD matrices, $\varphi(X) = -S(X)$

$$X^{k+1} = \frac{\exp(\log(X^k) - t_k \nabla f(X^k))}{\operatorname{tr}[\exp(\log(X^k) - t_k \nabla f(X^k))]}.$$

Relative smoothness

(Bauschke et al., 2017) and (Lu et al., 2018)

A function f is L-smooth relative to φ if for $L > 0$

$$L\varphi - f \quad \text{convex.}$$

A function f is μ-strongly convex relative to φ if for $\mu > 0$

$$f - \mu\varphi \quad \text{convex.}$$

Relative smoothness

(Bauschke et al., 2017) and (Lu et al., 2018)

A function f is L-smooth relative to φ if for $L > 0$

$$L\varphi - f \quad \text{convex.}$$

A function f is μ-strongly convex relative to φ if for $\mu > 0$

$$f - \mu\varphi \quad \text{convex.}$$

(Bauschke et al., 2017) and (Lu et al., 2018)

Mirror descent w/ $t_k = 1/L$ converges sublinearly $O(L/k)$ to global optimum if f is L-smooth relative to φ.

If, additionally, f is also μ-strongly convex relative to φ, then mirror descent will converge linearly $O((1 - \mu/L)^k)$ to global optimum.

Quantum Channel Capacities

For a linear map $\mathcal{N}$, define contraction coefficient as

$$C_{\mathcal{N}} = \sup_{X,Y\in\mathbb{H}^n_+} \left\{ \frac{S(\mathcal{N}(X)\,\|\,\mathcal{N}(Y))}{S(X\,\|\,Y)} : \operatorname{tr}[X] = \operatorname{tr}[Y] = 1,\; X \neq Y \right\},$$

and expansion coefficient as

$$E_{\mathcal{N}} = \inf_{X,Y\in\mathbb{H}^n_+} \left\{ \frac{S(\mathcal{N}(X)\,\|\,\mathcal{N}(Y))}{S(X\,\|\,Y)} : \operatorname{tr}[X] = \operatorname{tr}[Y] = 1,\; X \neq Y \right\},$$

Quantum Channel Capacities

For a linear map $\mathcal{N}$, define contraction coefficient as

$$C_{\mathcal{N}} = \sup_{X,Y\in\mathbb{H}^n_+} \left\{ \frac{S(\mathcal{N}(X) \| \mathcal{N}(Y))}{S(X \| Y)} : \operatorname{tr}[X] = \operatorname{tr}[Y] = 1, \ X \neq Y \right\},$$

and expansion coefficient as

$$E_{\mathcal{N}} = \inf_{X,Y\in\mathbb{H}^n_+} \left\{ \frac{S(\mathcal{N}(X) \| \mathcal{N}(Y))}{S(X \| Y)} : \operatorname{tr}[X] = \operatorname{tr}[Y] = 1, \ X \neq Y \right\},$$

Remark

Can interpret $C_{\mathcal{N}}$ and $E_{\mathcal{N}}$ as quantum relative entropy versions of min. and max. eigenvalues of $\mathcal{N}$

Also, $0 \leq E_{\mathcal{N}} \leq C_{\mathcal{N}} \leq 1$ follows (nontrivially) from joint convexity of S

Quantum Channel Capacities

For a linear map $\mathcal{N}$, define contraction coefficient as

$$C_{\mathcal{N}} = \sup_{X,Y\in\mathbb{H}^n_+} \left\{ \frac{S(\mathcal{N}(X)\,\|\,\mathcal{N}(Y))}{S(X\,\|\,Y)} : \operatorname{tr}[X] = \operatorname{tr}[Y] = 1,\ X \neq Y \right\},$$

and expansion coefficient as

$$E_{\mathcal{N}} = \inf_{X,Y\in\mathbb{H}^n_+} \left\{ \frac{S(\mathcal{N}(X)\,\|\,\mathcal{N}(Y))}{S(X\,\|\,Y)} : \operatorname{tr}[X] = \operatorname{tr}[Y] = 1,\ X \neq Y \right\},$$

Theorem 1 (HSF, 2023)

Recall quantum mutual information:

$$I(X) \coloneqq S(X) + S(\mathcal{N}(X)) - S(\mathcal{M}(X)).$$

Negative quantum mutual information is $(1 + C_{\mathcal{N}} - E_{\mathcal{M}})$-smooth and $(1 + E_{\mathcal{N}} - C_{\mathcal{M}})$-strongly convex rel. to $-S$.

Numerical results

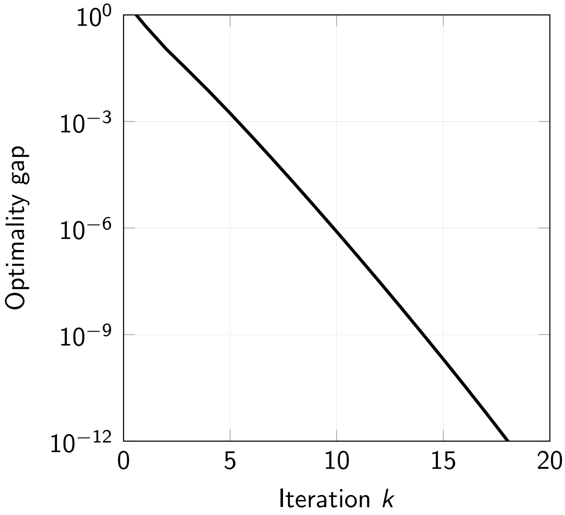

Figure: Quantum channel capacity over $X \in \mathbb{H}^{64}$.

Relationship to Blahut-Arimoto

Mirror descent applied to channel capacities is equivalent to seminal **Blahut-Arimoto** algorithm from information theory!

- Blahut-Arimoto algorithm first introduced (Blahut, 1972) and (Arimoto, 1972) to solve for classical channel capacities.
- Extended to quantum channel capacities in (Nagaoka, 1998), (Li & Cai, 2019), (Ramakrishnan et al., 2021).
- Derived using alternating optimization, but leads to same iterations (and very similar convergence criteria and rates).

Table of Contents

Constrained Channel Capacities

Constrained quantum channel capacity

$$\begin{aligned} \max_{X \in \mathbb{H}^n} \quad & I(X) \\ \text{subj. to} \quad & \langle A_i, X \rangle \leq b_i, \quad \forall i = 1, \ldots, p \\ & \text{tr}[X] = 1 \\ & X \succeq 0, \end{aligned}$$

where $A_i \in \mathbb{H}^n$ and $b_i \in \mathbb{R}$ encode linear constraints.

Constrained Channel Capacities

Constrained quantum channel capacity

$$\begin{aligned} \max_{X \in \mathbb{H}^n} \quad & I(X) \\ \text{subj. to} \quad & \langle A_i, X \rangle \leq b_i, \quad \forall i = 1, \ldots, p \\ & \text{tr}[X] = 1 \\ & X \succeq 0, \end{aligned}$$

where $A_i \in \mathbb{H}^n$ and $b_i \in \mathbb{R}$ encode linear constraints.

Not obvious how to perform mirror descent iteration now

$$\begin{aligned} X^{k+1} \quad = \quad \operatorname*{arg\,min}_{X \in \mathbb{H}^n} \quad & \langle \nabla I(X^k), X \rangle + \frac{1}{t_k} D_\varphi(X \,\|\, X^k) \\ \text{subj. to} \quad & \langle A_i, X \rangle \leq b_i, \quad \forall i = 1, \ldots, p \\ & \text{tr}[X] = 1 \\ & X \succeq 0 \end{aligned}$$

Primal-dual hybrid gradient

Consider linearly constrained convex optimization problem

$$\begin{aligned} \min_{x \in \mathcal{X}} \quad & f(x) \\ \text{subj. to} \quad & b - Ax \leq 0 \end{aligned}$$

Primal-dual hybrid gradient (PDHG) solves saddle point problem

$$\inf_{x \in \mathcal{X}} \sup_{z \geq 0} \quad \mathcal{L}(x, z) := f(x) + \langle z, Ax - b \rangle,$$

Primal-dual hybrid gradient

Consider linearly constrained convex optimization problem

$$\begin{aligned} \min_{x \in \mathcal{X}} \quad & f(x) \\ \text{subj. to} \quad & b - Ax \leq 0 \end{aligned}$$

Primal-dual hybrid gradient (PDHG) solves saddle point problem

$$\inf_{x \in \mathcal{X}} \sup_{z \geq 0} \quad \mathcal{L}(x, z) := f(x) + \langle z, Ax - b \rangle,$$

using alternating mirror descent steps on primal and dual variables

$$\begin{aligned} \bar{z}^{k+1} &= z^k + \theta_k (z^k - z^{k-1}) \\ x^{k+1} &= \arg\min_{x \in \mathcal{X}} \left\{ \langle \nabla f(x) + A^* \bar{z}^{k+1}, x \rangle + \frac{1}{\tau_k} D_\varphi(x \| x^k) \right\} \\ z^{k+1} &= \arg\min_{z \geq 0} \left\{ -\langle z, Ax^{k+1} - b \rangle + \frac{1}{2\gamma_k} \|z - z^k\|_2^2 \right\} \end{aligned}$$

Primal-dual hybrid gradient

Convergence result is simple extension of (Chambolle & Pock, 2016)

Observation

If f is L-**smooth relative to** φ, then PDHG with constant step sizes $\tau_k = \tau$ and $\gamma_k = \gamma$ satisfying

$$\left(\frac{1}{\tau} - L\right) D_\varphi(x \| x') + \frac{1}{2\gamma}\|z - z'\|_2^2 \geq \langle z - z', A(x - x')\rangle,$$

for all $x, x' \in \mathcal{X}$ and $z, z' \geq 0$, will have **ergodic sublinear convergence** to the primal-dual solution.

Primal-dual hybrid gradient

Convergence result is simple extension of (Chambolle & Pock, 2016)

Observation

If f is L**-smooth relative to** φ, then PDHG with constant step sizes $\tau_k = \tau$ and $\gamma_k = \gamma$ satisfying

$$\left(\frac{1}{\tau} - L\right) D_\varphi(x \| x') + \frac{1}{2\gamma}\|z - z'\|_2^2 \geq \langle z - z', A(x - x')\rangle,$$

for all $x, x' \in \mathcal{X}$ and $z, z' \geq 0$, will have **ergodic sublinear convergence** to the primal-dual solution.

Can also obtain ergodic sublinear convergence with **backtracking PDHG** using similar ideas as (Jiang & Vandenberghe, 2022)

Constrained Channel Capacities

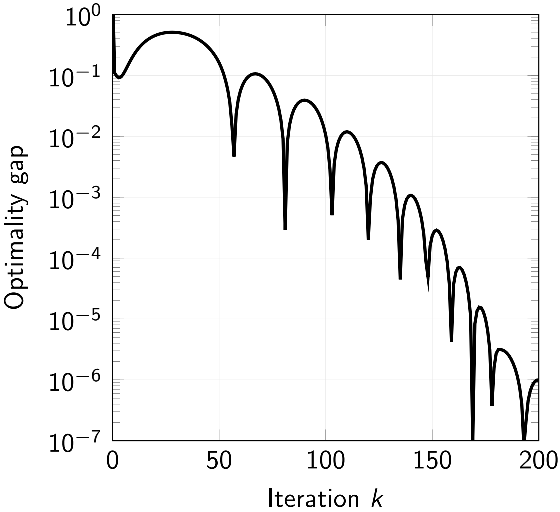

Figure: Quantum channel capacity over $X \in \mathbb{H}^{64}$ with 5 additional linear inequality constraints.

Conclusion

Summary:

- Mirror descent can efficiently solve problems in quantum information

Conclusion

Summary:

- Mirror descent can efficiently solve problems in quantum information
- Relative smoothness and strong convexity provide convergence guarantees

Conclusion

Summary:

- Mirror descent can efficiently solve problems in quantum information
- Relative smoothness and strong convexity provide convergence guarantees
- PDHG allows us to solve problems with more complicated constraints

Conclusion

Summary:

- Mirror descent can efficiently solve problems in quantum information
- Relative smoothness and strong convexity provide convergence guarantees
- PDHG allows us to solve problems with more complicated constraints
- Other applications:
  - Classical-quantum, quantum-quantum channel capacities
  - Quantum rate-distortion
  - Relative entropy of entanglement

Conclusion

Summary:

- Mirror descent can efficiently solve problems in quantum information
- Relative smoothness and strong convexity provide convergence guarantees
- PDHG allows us to solve problems with more complicated constraints
- Other applications:
  - Classical-quantum, quantum-quantum channel capacities
  - Quantum rate-distortion
  - Relative entropy of entanglement

Outlook:

- Ergodic linear convergence under relative strong convexity?
- Solve general quantum relative entropy programs using similar ideas?

Conclusion

Summary:

- Mirror descent can efficiently solve problems in quantum information
- Relative smoothness and strong convexity provide convergence guarantees
- PDHG allows us to solve problems with more complicated constraints
- Other applications:
  - Classical-quantum, quantum-quantum channel capacities
  - Quantum rate-distortion
  - Relative entropy of entanglement

Outlook:

- Ergodic linear convergence under relative strong convexity?
- Solve general quantum relative entropy programs using similar ideas?

K. He, J. Saunderson, H Fawzi, "A Bregman Proximal Perspective on Classical and Quantum Blahut-Arimoto Algorithms," IEEE Transactions on Information Theory, vol. 70, no. 8, pp. 5710-5730, Aug. 2024.