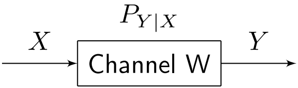
Efficient Approximation of Channel Capacities

David Sutter
Quantum Information Theory Group, ETH Zürich
www.qit.ethz.ch

Tobias Sutter, Peyman Mohajerin Esfahani, John Lygeros
Automatic Control Laboratory, ETH Zürich
WWW.CONTROL.ETHZ.CH

Channel Capacity

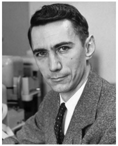


- Discrete memoryless channel (DMC)
  $\mathsf{W}_{x,y} := W(y|x) = \mathbb{P}\big(Y = y | X = x\big)$, where $\mathcal{X} = \{1, 2, \ldots, N\}$, $\mathcal{Y} = \{1, 2, \ldots, M\}$

Definition ([Shannon'48])

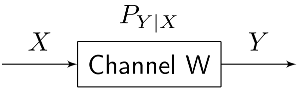
The capacity of a DMC channel is defined as

$$C := \max_{p \in \Delta_N} I(p, \mathsf{W})$$



- Operational meaning to the definition of the capacity C as the maximum number of bits we can transmit reliably over the channel W
- Mutual information $I(p, \mathsf{W}) := \sum_{i=1}^{N} p_i \sum_{k=1}^{M} \mathsf{W}_{i,k} \log \left(\frac{\mathsf{W}_{i,k}}{\sum_{\ell=1}^{N} p_\ell \mathsf{W}_{\ell,k}} \right)$

Channel Capacity (cont'd)



- Capacity of W under additional input cost constraint $\mathbb{E}[s(X)] \leqslant S$

$$\mathsf{P}: \qquad C_S = \begin{cases} \max\limits_{p} & I(p, \mathsf{W}) \\ \text{s.t.} & s^\mathsf{T} p \leqslant S \\ & p \in \Delta_N. \end{cases}$$

  - Mutual information $I(p, \mathsf{W}) = \sum_{i=1}^{N} p_i \sum_{k=1}^{M} \mathsf{W}_{i,k} \log\left(\frac{\mathsf{W}_{i,k}}{\sum_{\ell=1}^{N} p_\ell \mathsf{W}_{\ell,k}}\right)$
  - finite-dimensional convex optimization problem
  - non-smooth objective function

State of the Art

- **Blahut-Arimoto Algorithm**: [Blahut'72] and [Arimoto'72]
  - Iterative method based on the primal problem
  - Specifically tuned for this problem
  - Difficulties for large input alphabet sizes N

State of the Art

- **Blahut-Arimoto Algorithm**: [Blahut'72] and [Arimoto'72]
  - Iterative method based on the primal problem
  - Specifically tuned for this problem
  - Difficulties for large input alphabet sizes N
- **Geometric programming and duality**: [Mung & Boyd'04]
  - Showed that the dual program can be recast as a geometric program
  - Generate analytical upper and lower bounds on the capacity problem
  - Point out connection between the dual programs of the capacity and the rate distortion problem

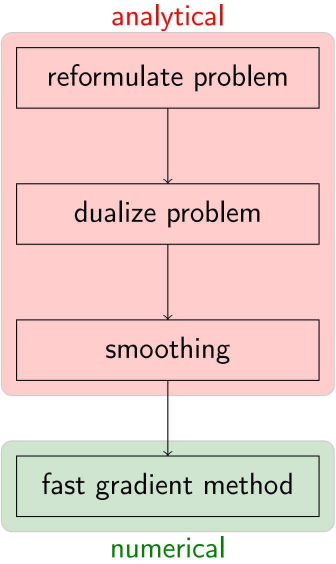
State of the Art

- **Blahut-Arimoto Algorithm**: [Blahut'72] and [Arimoto'72]
  - Iterative method based on the primal problem
  - Specifically tuned for this problem
  - Difficulties for large input alphabet sizes N
- **Geometric programming and duality**: [Mung & Boyd'04]
  - Showed that the dual program can be recast as a geometric program
  - Generate analytical upper and lower bounds on the capacity problem
  - Point out connection between the dual programs of the capacity and the rate distortion problem
- **Cutting plane algorithm**: [Huang & Meyn'05]
  - Iteratively approximate the mutual information by linear functionals
  - Solve an LP in each iteration step

State of the Art

- **Blahut-Arimoto Algorithm**: [Blahut'72] and [Arimoto'72]
  - Iterative method based on the primal problem
  - Specifically tuned for this problem
  - Difficulties for large input alphabet sizes N
- **Geometric programming and duality**: [Mung & Boyd'04]
  - Showed that the dual program can be recast as a geometric program
  - Generate analytical upper and lower bounds on the capacity problem
  - Point out connection between the dual programs of the capacity and the rate distortion problem
- **Cutting plane algorithm**: [Huang & Meyn'05]
  - Iteratively approximate the mutual information by linear functionals
  - Solve an LP in each iteration step
- **Many more**: Lapidoth, Ben-Tal, Lasserre, . . .

Outline



introduce additional decision variable

closed form Lagrange dual function

entropy maximization

a priori & a posteriori error

Equivalent Primal Reformulation

Idea: Introduce the output distribution $q \in \Delta_M$ as additional decision variable

Assumption (w.l.o.g.): $S \leqslant S_{\max}$

Lemma

The channel capacity problem P *is equivalent to*

$$C_S = \begin{cases} \max\limits_{p,q} & -r^\mathsf{T} p + H(q) \\ \text{s.t.} & \mathsf{W}^\mathsf{T} p = q \\ & s^\mathsf{T} p = S \\ & p \in \Delta_N, \ q \in \Delta_M. \end{cases}$$

$r_i := -\sum_{j=1}^{M} \mathsf{W}_{ij} \log(\mathsf{W}_{ij})$

Dual Capacity Problem

Its dual program (with strong duality) is

$$\mathsf{D}: \quad C_S = \min_{\lambda} \left\{ G(\lambda) + F(\lambda) \ : \ \lambda \in \mathbb{R}^M \right\},$$

with the Lagrange dual function given by

$$G(\lambda) = \begin{cases} \max\limits_{p} & -r^{\mathsf{T}} p + \lambda^{\mathsf{T}} \mathsf{W}^{\mathsf{T}} p \\ \text{s.t.} & s^{\mathsf{T}} p = S \\ & p \in \Delta_N \end{cases} \quad \text{and} \quad F(\lambda) = \begin{cases} \max\limits_{q} & H(q) - \lambda^{\mathsf{T}} q \\ \text{s.t.} & q \in \Delta_M \end{cases}$$

Dual Capacity Problem

Its dual program (with strong duality) is

$$\text{D}: \quad C_S = \min_{\lambda} \left\{ G(\lambda) + F(\lambda) \;:\; \lambda \in \mathbb{R}^M \right\},$$

non-compact

with the Lagrange dual function given by

$$G(\lambda) = \begin{cases} \max\limits_{p} & -r^{\mathsf{T}} p + \lambda^{\mathsf{T}} \mathsf{W}^{\mathsf{T}} p \\ \text{s.t.} & s^{\mathsf{T}} p = S \\ & p \in \Delta_N \end{cases} \quad \text{and} \quad F(\lambda) = \begin{cases} \max\limits_{q} & H(q) - \lambda^{\mathsf{T}} q \\ \text{s.t.} & q \in \Delta_M \end{cases}$$

non-smooth

smooth, entropy maximization

$$F(\lambda) = \log\left(\sum_{i=1}^{M} 2^{-\lambda_i} \right)$$

Bounding Dual Variables

Assumption (Non-singular channel matrix W)

$\gamma := \min_{i,j} \mathsf{W}_{ij} > 0$

Since the mutual information is continuous in W, in case of a singular entry one could perturb this entry.

Bounding Dual Variables

Assumption (Non-singular channel matrix W)

$\gamma := \min_{i,j} \mathsf{W}_{ij} > 0$

Since the mutual information is continuous in W, in case of a singular entry one could perturb this entry.

Proposition

Under the above assumption, the dual program D *is equivalent to*

$$C_S = \min_{\lambda} \{G(\lambda) + F(\lambda) \ : \ \lambda \in Q\},$$

where $Q := \left\{\lambda \in \mathbb{R}^M \ : \ \|\lambda\|_2 \leqslant M \log(\gamma^{-1})\right\}$

Smoothing Step

Consider a smoothing parameter $\nu > 0$ and the program

$$G_\nu(\lambda) = \begin{cases} \max\limits_p & \lambda^\mathsf{T} \mathsf{W}^\mathsf{T} p - r^\mathsf{T} p + \nu H(p) - \nu \log(N) \\ \text{s.t.} & s^\mathsf{T} p = S \\ & p \in \Delta_N, \end{cases}$$

which is a modified entropy maximization and has the solution [Cover'06]

$$p_\nu(\lambda)_i = 2^{\mu_1 + \frac{1}{\nu}(\lambda^\mathsf{T} \mathsf{W}^\mathsf{T} - r^\mathsf{T})_i + \mu_2 s_i},$$

where $\mu_1, \mu_2 \in \mathbb{R}$ are such that $s^\mathsf{T} p_\nu(\lambda) = S$ and $p_\nu(\lambda) \in \Delta_N$.

Uniform Approximation: $G_\nu(\lambda) \leqslant G(\lambda) \leqslant G_\nu(\lambda) + \nu \log(N)$

Smoothing Step

Consider a smoothing parameter $\nu > 0$ and the program

$$G_\nu(\lambda) = \begin{cases} \max\limits_{p} & \lambda^\mathsf{T} \mathsf{W}^\mathsf{T} p - r^\mathsf{T} p + \nu H(p) - \nu \log(N) \\ \text{s.t.} & s^\mathsf{T} p = S \\ & p \in \Delta_N, \end{cases}$$

which is a modified entropy maximization and has the solution [Cover'06]

$$p_\nu(\lambda)_i = 2^{\mu_1 + \frac{1}{\nu}(\lambda^\mathsf{T} \mathsf{W}^\mathsf{T} - r^\mathsf{T})_i + \mu_2 s_i},$$

where $\mu_1, \mu_2 \in \mathbb{R}$ are such that $s^\mathsf{T} p_\nu(\lambda) = S$ and $p_\nu(\lambda) \in \Delta_N$.

Uniform Approximation: $G_\nu(\lambda) \leqslant G(\lambda) \leqslant G_\nu(\lambda) + \nu \log(N)$

Lemma ([Nesterov'05])

The gradient $\nabla G_\nu(\lambda) = \mathsf{W}^\mathsf{T} p_\nu(\lambda)$ *is Lipschitz continuous with constant* $L_\nu = \frac{1}{\nu}$.

Smoothing Step (cont'd)

Consider the smooth, convex optimization problem over a compact set

$$\mathsf{D}_\nu : \begin{cases} \min\limits_{\lambda} & F(\lambda) + G_\nu(\lambda) \\ \text{s.t.} & \lambda \in Q, \end{cases}$$

Smoothing Step (cont'd)

Consider the smooth, convex optimization problem over a compact set

$$\mathsf{D}_\nu : \begin{cases} \min\limits_{\lambda} & F(\lambda) + G_\nu(\lambda) \\ \text{s.t.} & \lambda \in Q, \end{cases}$$

π_Q is the projection operator of Q

Algorithm ($\diamond$): Optimal scheme for smooth optimization [Nesterov'05]

For $k \geqslant 0$ **do**

- **Step 1:** Compute $\nabla F(\lambda_k) + \nabla G_\nu(\lambda_k)$
- **Step 2:** $y_k = \pi_Q\left(-\frac{1}{L_\nu}\left(\nabla F(\lambda_k) + \nabla G_\nu(\lambda_k)\right) + \lambda_k\right)$
- **Step 3:** $z_k = \pi_Q\left(-\frac{1}{L_\nu}\sum_{i=0}^{k}\frac{i+1}{2}\left(\nabla F(\lambda_i) + \nabla G_\nu(\lambda_i)\right)\right)$
- **Step 4:** $\lambda_{k+1} = \frac{2}{k+3} z_k + \frac{k+1}{k+3} y_k$

Error Bound

Theorem ([Nesterov'05])

Consider the parameter

$$D_1 := \tfrac{1}{2}(M\log(\gamma^{-1}))^2, \quad D_2 := \log(N), \quad \nu = \frac{2}{n+1}\sqrt{\frac{D_1}{D_2}}.$$

Then after n iterations of Algorithm $(\diamond)$ *we can generate*

$$\hat{\lambda} = y_n \in Q, \qquad \hat{p} = \sum_{i=0}^{n} \frac{2(i+1)}{(n+1)(n+2)} p_\nu(\lambda_i) \in \Delta_N.$$

Then, $\quad 0 \leqslant F(\hat{\lambda}) + G(\hat{\lambda}) - I(\hat{p}, W) \leqslant \frac{4}{n+1}\sqrt{D_1 D_2} + \frac{4D_1}{(n+1)^2}$

a posteriori error

a priori error

- $C_{\mathsf{UB}} := F(\hat{\lambda}) + G(\hat{\lambda}),$ $\qquad C_{\mathsf{LB}} := I(\hat{p}, \mathsf{W})$

Computational Complexity

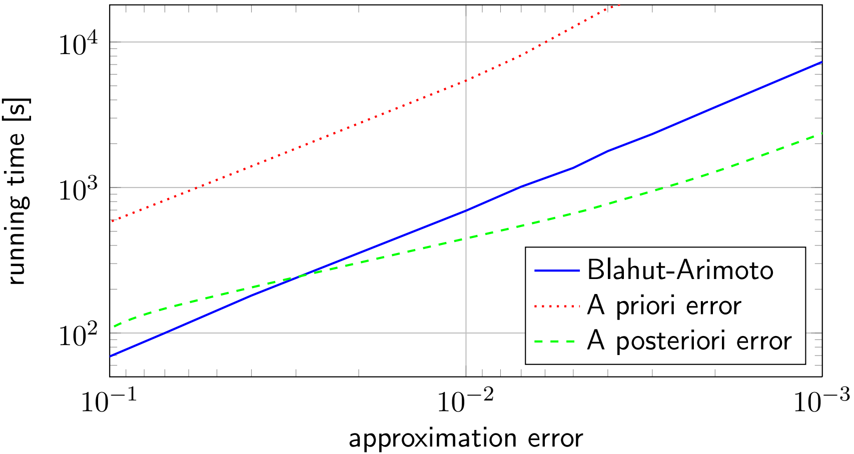
N = input alphabet size
M = output alphabet size

	Blahut-Arimoto	New method
Complexity of one iteration step	$O(MN^2)$	$O(MN)$
Complexity for an ε-solution	$O\left(\frac{MN^2 \log N}{\varepsilon}\right)$	$O\left(\frac{M^2 N\sqrt{\log N}}{\varepsilon}\right)$

ETH Zürich

Simulation Results — Example 1

Consider a DMC having a channel matrix $\mathsf{W} \in \mathbb{R}^{N \times M}$ with $N = 10000$ and $M = 100$, whose entries are chosen i.i.d. uniformly distributed in $[0, 1]$.



Simulation Results — Example 2

Consider a DMC having a channel matrix $\mathsf{W} \in \mathbb{R}^{N \times M}$ with $N = 10^5$ and $M = 10$, whose entries are chosen i.i.d. uniformly distributed in $[0, 1]$.

Table : Example 2

A priori error	1	0.1	0.01	0.001
$C_{\mathsf{UB}}(W)$	1.0938	1.0543	1.0516	1.0514
$C_{\mathsf{LB}}(W)$	1.0101	1.0512	1.0514	1.0514
A posteriori error	0.0837	0.0031	$2.5 \cdot 10^{-4}$	$2.9 \cdot 10^{-5}$
Time [s]	23	227	1833	17 529
Iterations	1249	12 329	123 132	1 231 160

Blahut-Arimoto algorithm: single iteration ↯

Continuous Input — Countable Output

- Input alphabet $\mathcal{X} \subseteq \mathbb{R}$, output alphabet $\mathcal{Y} = \mathbb{N}_0$ (e.g. discrete-time Poisson channel)
- Peak-power constraint: $\mathbb{A} \subseteq \mathcal{X}$ compact, $\mathbb{P}\big(X \in \mathbb{A}\big) = 1$
- Average-power constraint: s continuous function
- Channel capacity

$$C_{\mathbb{A},S}(\mathsf{W}) = \left\{ \begin{array}{ll} \sup\limits_{p} & I(p, W) \\ \text{s. t.} & \mathbb{E}[s(X)] \leqslant S \\ & p \in \mathcal{D}(\mathbb{A}), \end{array} \right.$$

$\uparrow$

space of probability densities on $\mathbb{A}$

Truncation

$$W_M(i|x) := \begin{cases} W(i|x) + \frac{1}{M}\sum_{j\geqslant M} W(j|x), & i \in \{0,1,\ldots,M-1\} \\ 0, & i \geqslant M. \end{cases}$$

Channel W_M has input alphabet $\mathcal{X}$ and output alphabet $\{0,1,\ldots,M-1\}$

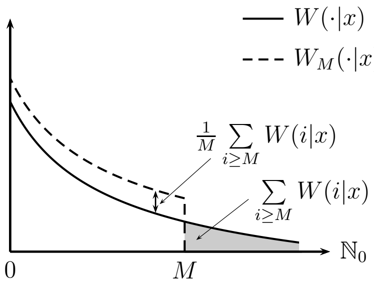


Figure : Pictorial representation of the M-truncated channel counterpart

Truncation

$$W_M(i|x) := \begin{cases} W(i|x) + \frac{1}{M} \sum_{j \geqslant M} W(j|x), & i \in \{0, 1, \ldots, M-1\} \\ 0, & i \geqslant M. \end{cases}$$

Channel W_M has input alphabet $\mathcal{X}$ and output alphabet $\{0, 1, \ldots, M-1\}$

Assumption (Tail decay)

For $M \in \mathbb{N}_0$ and $k \in (0, 1)$

$$R_k(M) := \sum_{i \geqslant M} \big(\sup_{x \in \mathcal{X}} W(i|x) \big)^k < \infty$$

Theorem (Truncation error)

Under the above assumption

$$\left| C_{\mathbb{A},S}(W) - C_{\mathbb{A},S}(W_M) \right| \leqslant \frac{2 \log(e)}{e(1-k)} \left[M^{1-k} \big(R_1(M) \big)^k + R_k(M) \right]$$

Primal Program

- Introduce a linear operator $\mathcal{W}$ and its adjoint $\mathcal{W}^\star$

$$\mathcal{W} : \mathbb{R}^M \to \mathrm{L}^\infty(\mathbb{A}), \qquad \mathcal{W}\lambda(x) := \sum_{i=1}^{M} W_M(i-1|x)\lambda_i$$

$$\mathcal{W}^\star : \mathrm{L}^1(\mathbb{A}) \to \mathbb{R}^M, \qquad (\mathcal{W}^\star p)_i := \int_X W_M(i-1|x)p(x)\,\mathrm{d}x$$

- Channel capacity of truncated channel

$$\mathsf{P}: \quad C_{\mathbb{A},S}(W_M) = \begin{cases} \sup\limits_{p,q} & -\langle p, r\rangle + H(q) \\ \text{s. t.} & \mathcal{W}^\star p = q \\ & \langle p, s\rangle = S \\ & p \in \mathcal{D}(\mathbb{A}),\ q \in \Delta_M, \end{cases}$$

Dual Program

Its dual program (with strong duality) is

$$\text{D}: \quad C_{\mathbb{A},S}(W_M) = \min_{\lambda} \left\{ G(\lambda) + F(\lambda) \ : \ \lambda \in \mathbb{R}^M \right\},$$

non-compact

with the Lagrange dual function given by

$$G(\lambda) = \begin{cases} \sup\limits_{p} & \langle p, \mathcal{W}\lambda \rangle - \langle p, r \rangle \\ \text{s.t.} & \langle p, s \rangle = S \\ & p \in \mathcal{D}(\mathbb{A}) \end{cases} \quad \text{and} \quad F(\lambda) = \begin{cases} \max\limits_{q} & H(q) - \lambda^{\mathsf{T}} q \\ \text{s.t.} & q \in \Delta_M . \end{cases}$$

non-smooth

infinite dimensional

smooth, entropy maximization

$$F(\lambda) = \log \left(\textstyle\sum_{i=1}^{M} 2^{-\lambda_i} \right)$$

Bounding Dual Variables

Assumption (Non-singular channel W_M)

$$\gamma_M := \min_{y \in \{0,1,\ldots,M-1\}} \min_{x \in \mathbb{A}} W_M(y|x) > 0$$

In case $\sum_{j \geqslant M} W(j|x) > 0$ for all x, a lower bound can be given by $\gamma_M \geqslant \frac{1}{M} \min_x \sum_{j \geqslant M} W(j|x)$.

Bounding Dual Variables

Assumption (Non-singular channel W_M)

$$\gamma_M := \min_{y \in \{0,1,\ldots,M-1\}} \min_{x \in \mathbb{A}} W_M(y|x) > 0$$

In case $\sum_{j \geqslant M} W(j|x) > 0$ for all x, a lower bound can be given by $\gamma_M \geqslant \frac{1}{M} \min_x \sum_{j \geqslant M} W(j|x)$.

Proposition

Under the above assumption, the dual program D *is equivalent to*

$$C_{\mathbb{A},S}(W_M) = \min_{\lambda} \{G(\lambda) + F(\lambda) \ : \ \lambda \in Q\},$$

where $Q := \{\lambda \in \mathbb{R}^M \ : \ \|\lambda\|_2 \leqslant M \log(\gamma_M^{-1})\}$

Smoothing Step

For $\nu > 0$

$$G_\nu(\lambda) = \begin{cases} \sup\limits_{p} & \langle p, \mathcal{W}\lambda\rangle - \langle p, r\rangle + \nu h(p) - \nu \log(\rho) \\ \text{s.t.} & \langle p, s\rangle = S \\ & p \in \mathcal{D}(\mathbb{A}), \end{cases}$$

which is a modified entropy maximization and has the solution [Cover'06]

$$p_\nu^\lambda(x) = 2^{\mu_1 + \frac{1}{\nu}(\mathcal{W}\lambda(x) - r(x)) + \mu_2 s(x)}, \quad x \in \mathbb{A},$$

where $\mu_1, \mu_2 \in \mathbb{R}$ are such that $\langle p_\nu^\lambda, s\rangle = S$ and $p_\nu^\lambda \in \mathcal{D}(\mathbb{A})$.

Uniform Approximation: $G_\nu(\lambda) \leqslant G(\lambda) \leqslant G_\nu(\lambda) + \underbrace{\iota(\nu)}_{\lim_{\nu\to 0} \iota(\nu) = 0}$

Smoothing Step

For $\nu > 0$

$$G_\nu(\lambda) = \begin{cases} \sup\limits_{p} & \langle p, \mathcal{W}\lambda \rangle - \langle p, r \rangle + \nu h(p) - \nu \log(\rho) \\ \text{s.t.} & \langle p, s \rangle = S \\ & p \in \mathcal{D}(\mathbb{A}), \end{cases}$$

which is a modified entropy maximization and has the solution [Cover'06]

$$p_\nu^\lambda(x) = 2^{\mu_1 + \frac{1}{\nu}(\mathcal{W}\lambda(x) - r(x)) + \mu_2 s(x)}, \quad x \in \mathbb{A},$$

where $\mu_1, \mu_2 \in \mathbb{R}$ are such that $\langle p_\nu^\lambda, s \rangle = S$ and $p_\nu^\lambda \in \mathcal{D}(\mathbb{A})$.

Uniform Approximation: $G_\nu(\lambda) \leqslant G(\lambda) \leqslant G_\nu(\lambda) + \underbrace{\iota(\nu)}_{\lim_{\nu \to 0} \iota(\nu) = 0}$

Lemma

The gradient $\nabla G_\nu(\lambda) = \mathcal{W}^\star p_\nu^\lambda$ is Lipschitz continuous with constant $L_\nu = \frac{1}{\nu}$.

Error Bound

Theorem

Consider the parameter $\alpha = \dagger$, $\nu = \frac{\varepsilon/\alpha}{\log(\alpha/\varepsilon)}$ *and* $n \geqslant \frac{1}{\varepsilon}\sqrt{8D_1\alpha}\sqrt{\log(\varepsilon^{-1}) + \log(\alpha) + \frac{1}{4}}$. *Then after* n *iterations of Algorithm* $(\diamond)$ *we can generate*

$$\hat{\lambda} = y_n \in Q \qquad \text{and} \qquad \hat{p} = \sum_{k=0}^{n} \frac{2(i+1)}{(n+1)(n+2)} p_\nu^{x_k} \in \mathcal{D}(\mathbb{A}).$$

$\Rightarrow$

$$0 \leqslant F(\hat{\lambda}) + G(\hat{\lambda}) - I(\hat{p}, W) \leqslant \varepsilon$$

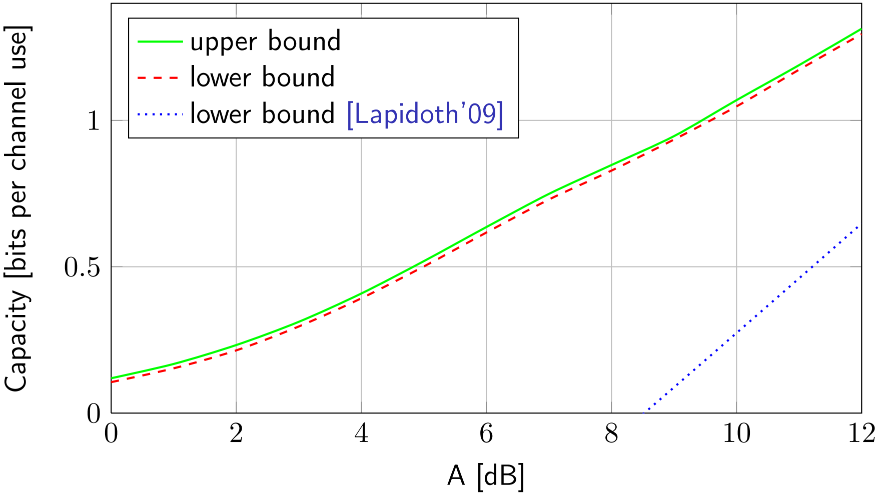
a posteriori error   a priori error

- $C_{\mathrm{UB}} := F(\hat{\lambda}) + G(\hat{\lambda})$, $C_{\mathrm{LB}} := I(\hat{p}, \mathrm{W})$

Discrete-Time Poisson Channel

- $W(y|x) = e^{-(x+\eta)} \frac{(x+\eta)^y}{y!}, \quad y \in \mathbb{N}_0,\, x \in \mathbb{R}_{\geqslant 0}$
- Important example to model optical communication systems [Shamai'90]
- Peak-power constraint $\mathbb{P}\big(X \in [0, A]\big) = 1$
- No analytic expression for the capacity of the Poisson channel with a peak-power constraint is known
  - Analytical lower bound is available
- Channel has fast decaying tail $\Rightarrow$ output alphabet truncation possible

Discrete-Time Poisson Channel (cont'd)



Outlook

arXiv: soon ☺

- Technical report
- Capacity of a classical-quantum channel
  - More precise modelling of optical channels should include quantum-mechanical effects
- Capacity of a random channel
- Approximate dynamic programming
  - Topic of my next coffee talk? :)

Acknowledgements and Questions

Special thanks to

- Stefan Richter
- Prof. Renato Renner

Acknowledgements and Questions

Questions?

Entropy Maximization

Consider the optimization problem

$$\left\{ \begin{array}{ll} \sup\limits_{p} & h(p) + \langle p, c \rangle \\ \text{s.t.} & \langle p, s \rangle = S \\ & p \in \mathcal{D}(\mathbb{A}), \end{array} \right. \tag{1}$$

with $c, s \in \mathrm{L}^\infty(\mathbb{A})$.

Lemma ([Boltzmann, 1877])

Let $p^\star_\mu(x) = 2^{\mu_1 + c(x) + \mu_2 s(x)}$, where $\mu_1, \mu_2 \in \mathbb{R}$ are chosen such that $p^\star_\mu$ satisfies the constraints in (1). Then $p^\star_\mu$ uniquely solves (1).

- straightforward extension to multiple moment contraints
- find μ_i using semidefinite programming [Lasserre, 2009]

Capacity of a Classical-Quantum (CQ) Channel

- $\mathcal{D}(\mathcal{H}) := \{\rho \in \mathcal{H} |\, \mathsf{tr}[\rho] = 1, \rho \geqslant 0\}$ is the set of density operators on a Hilbert space $\mathcal{H}$
- CQ-channel $W : \mathcal{X} \to \mathcal{D}(\mathcal{H})$, $x \longmapsto \rho_x$
- von Neumann entropy $H(\rho_x) := -\,\mathsf{tr}[\rho_x \log \rho_x]$
- Capacity of a CQ-channel

$$C_{\mathsf{cq}} = \begin{cases} \max\limits_{P_X} & H\left(\sum_{x\in\mathcal{X}} P_X(x)\rho_x\right) - \sum_{x\in\mathcal{X}} P_X(x) H(\rho_x) \\ \text{s.t.} & \sum_{x\in\mathcal{X}} s(x) P_X(x) \leqslant S \\ & P_X \in \Delta_N . \end{cases}$$