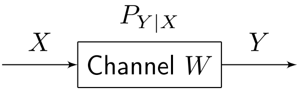
Capacity Approximation of Memoryless Channels with Countable Output Alphabets

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

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

Channel Capacity

- Memoryless channel with
  - continuous input alphabet $\mathcal{X} \subseteq \mathbb{R}$
  - countable output alphabet $\mathcal{Y} = \mathbb{N}_0$
  - transition density $W(y|x) := \mathbb{P}\big(Y = y | X = x\big)$



- Constraints on the input
  - Peak-power constraint: $\mathbb{A} \subseteq \mathcal{X}$ compact, $\mathbb{P}\big(X \in \mathbb{A}\big) = 1$
  - Average-power constraint: $\mathbb{E}[s(X)] \leqslant S$ for $s \in \mathrm{L}^\infty(\mathbb{R})$

Capacity Formula

The capacity of such a channel is given by

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

- $\mathcal{D}(\mathbb{A})$ denotes the space of probability densities on $\mathbb{A}$

Channel Capacity (cont'd)

- Primal optimization problem

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

  - infinte-dimensional convex optimization problem
  - non-smooth objective function

- Approximation methods
  - Mostly restricted to the finite-dimensional case
    - Blahut-Arimoto algorihm [Blahut'72] and [Arimoto'72]
    - Geometric programming [Mung & Boyd'04]
  - Cutting plane algorithm: [Huang & Meyn'05]
    - Iteratively approximate the mutual information by linear functionals
    - Solve an infinite-dimensional LP in each iteration step

Outline

analytical

tail truncation	finite output alphabet
reformulate problem	introduce additional decision variable
dualize problem	closed form Lagrange dual function
smoothing	entropy maximization
fast gradient method	a priori & a posteriori error

numerical

Truncation

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

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

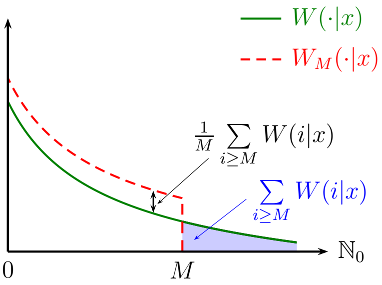


Figure : Pictorial representation of the M-truncated channel counterpart

Truncation (cont'd)

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

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

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

Equivalent Primal Reformulation

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

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

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

Lemma (Additional decision variable)

For $S \leqslant S_{\max}$ the capacity of truncated channel is given by

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

- $r(\cdot) := -\sum_{j=0}^{M-1} W_M(j|\cdot)\log(W_M(j|\cdot))$

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

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

Entropy Maximization

Consider the optimization problem

$$\begin{cases} \sup_{p} & h(p) + \langle p, c \rangle \\ \text{s.t.} & \langle p, s \rangle = S \\ & p \in \mathcal{D}(\mathbb{A}), \end{cases} \tag{1}$$

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]

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*

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

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

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

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

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

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

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* $L_{\nu} = \frac{1}{\nu}$.

Smooth Optimization

Consider the smooth, convex optimization problem over a compact set

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

Smooth Optimization

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 on 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

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

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

$\Rightarrow$

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

a posteriori error

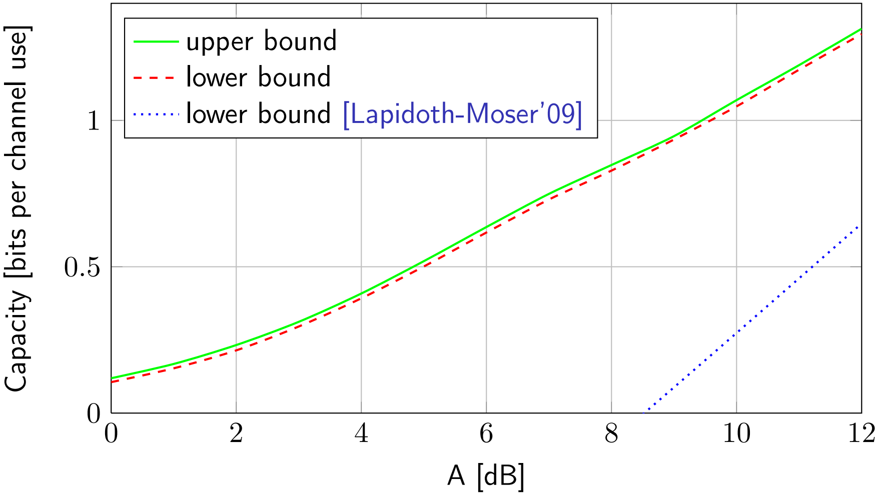
a priori error

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

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
- 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)



Thank you

Questions ?