Proposal: Convex Relaxations of Entropy-Constrained Compression

Marc Nasser
Department of Electrical Engineering
Stanford University

October 31, 2025

Problem

Most modern lossy compression systems try to balance distortion against rate to determine the level of compression they would like to achieve for their application. This trade-off can be formalized by entropy-constrained compression optimization: minimizing the entropy of compressed representation Y ($H(Y)$) while keeping expected distortion ($\mathbb{E}[d(X, Y)]$) below a specific target.

This problem is often solved by optimizing a rate-distortion objective with non-convex algorithms such as Lloyd-Max, or learned models. The issue is that they lack global optimality guarantees and do not have an explicit mechanism to impose fidelity on specific subsets of the data, where the subsets vary in importance.

However, many applications in practice require mixed-fidelity compression, where presentation of certain aspects of the data (such as metadata or crucial safety-critical information) are more important than others. The certification of such guarantees are hard whose empirical entropy closely approaches the theoretical convex lower bound on achievable rate. to be achieved with non-convex optimization.

Therefore, we aim to provide a convex formulation/relaxation of entropy-constrained compression that directly addresses these limitations. The convex perspective allows us to get globally optimal stochastic encoders (for a fixed reconstruction alphabet), exposes dual variables quantifying the marginal "price" of distortion and fidelity constraints, and provides auditable lower bounds on the best achievable rate. By solving sequential convex problems by iteratively constraining the encoder towards a deterministic function and simplifying the codebook, we obtain deployable compressors whose empirical entropy closely approaches the theoretical convex lower-bound on achievable rate, and allow us to estimate better the limits of our encoders.

Readings

1. R. Blahut, "Computation of Channel Capacity and Rate–Distortion Functions," *IEEE Transactions on Information Theory*, 1972.
2. S. Arimoto, "An Algorithm for Calculating the Capacity of Arbitrary Discrete Memoryless Channels," *IEEE Transactions on Information Theory*, 1972.
3. G. Peyré and M. Cuturi, *Computational Optimal Transport*, Foundations and Trends in Machine Learning, 2019.

Method

We formulate the compression problem as a convex optimization over stochastic encoders $Q \in \mathbb{R}^{m\times n}$, where $Q_{j|x} = P(Y = y_j \mid X = x)$ denotes the probability of mapping input x to reconstruction y_j. Given an empirical source distribution p_X and distortion measure $d(x, y)$, we seek an encoder minimizing mutual information subject to distortion constraints.

Since $H(Y)$ is nonconvex, we minimize the convex surrogate $I(X; Y)$ and progressively anneal it to approach $H(Y)$ as the encoder becomes deterministic. The optimization problem is

$$
\begin{aligned}
\min_{Q\geq 0} \quad & I(X;Y) = \sum_{x\in\mathcal{X}} p_X(x) \sum_{j=1}^{m} Q_{j|x} \log \frac{Q_{j|x}}{p_Y(j)} \\
\text{s.t.} \quad & \sum_{j=1}^{m} Q_{j|x} = 1, \quad \forall x, \\
& \sum_{x,j} p_X(x) Q_{j|x} d(x, y_j) \leq D, \\
& \sum_{j} Q_{j|x} d(x, y_j) \leq D', \quad \forall x \in \mathcal{S}, \\
& p_Y(j) = \sum_{x} p_X(x) Q_{j|x}.
\end{aligned}
$$

Here, $\mathcal{S}$ denotes a fidelity subset with tighter distortion tolerance $D' \ll D$. The objective $I(X;Y)$ is convex in Q, and the feasible set is affine, ensuring global optimality. The rate–distortion function $R(D)$ is convex and differentiable, with dual variable $\lambda^\star = -R'(D)$ quantifying the marginal trade-off.

After solving the convex program for a fixed codebook $\mathcal{Y} = \{y_j\}_{j=1}^m$, we refine $\mathcal{Y}$ via gradient updates $y_j \leftarrow y_j - \eta \nabla_{y_j} \mathbb{E}[d(X, y_j)]$, alternating between encoder and codebook updates until convergence.

To align the convex information measure $I(X;Y)$ with the entropy $H(Y)$, we employ deterministic annealing on Q. Introducing a temperature $\tau > 0$, we define $Q^{(\tau)}_{j|x}$, where we solve the convex problem for the current temperature τ, then gradually decrease τ and re-solve, reinitializing with the previous solution. As $\tau \downarrow 0$, the encoder becomes nearly deterministic, $H(Y|X;\tau) \to 0$, and $I(X;Y;\tau) \to H(Y)$. To further reduce entropy, low-mass components $(p_Y(j) < \varepsilon)$ are pruned or merged, preserving total probability and ensuring nonincreasing $H(Y)$.

All optimization components are implemented in `CVXPY`, which provides convex modeling primitives and solvers for relative-entropy programs.

Evaluation

We will evaluate performance using both qualitative and quantitative criteria. Qualitatively, we will visualize rate–distortion (R–D) curves, plot $I(X;Y)$, $H(Y)$, and $H(Y|X)$ across annealing stages, and examine the sensitivity of the solution to constraint variations. Quantitatively, we will report the achieved rate, average distortion, and duality gaps, and compare results with baselines such as entropy-constrained Lloyd quantization. Essentially, we will showcase the benefits of having convex tools as part of our analysis to make informed decisions.

Timeline

Week 1: Implement the convex formulation in `CVXPY` and verify correctness on synthetic data; generate initial R–D curves.
Week 2: Add fidelity constraints and analyze dual variables; evaluate sensitivity to distortion bounds. Integrate codebook updates and support refinement; compare to fixed-support baselines.
Week 3: Implement deterministic annealing schedule and pruning; visualize $I(X;Y)$, $H(Y)$, and $H(Y|X)$ evolution.

Week 4: Conduct ablations, statistical comparisons, and finalize report with plots and convex analysis.
Week 5: Prepare report and presentation