Project Milestone:
Convex Relaxations for Entropy-Constrained Compression

Marc Moussa Nasser

Introduction

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

This problem is often solved by optimizing a rate–distortion objective with nonconvex algorithms such as Lloyd–Max quantization or learned neural models. These approaches lack global optimality guarantees and do not offer an explicit mechanism to impose higher fidelity on specific subsets of the data whose importance varies across the source.

However, many practical applications require mixed-fidelity compression, where accurate reconstruction of certain aspects of the data (such as metadata or safety-critical information) is more important than others. With standard nonconvex optimization, it is difficult to certify that a practical encoder's empirical entropy is close to the theoretical convex lower bound on the achievable rate under such heterogeneous fidelity constraints.

Therefore, we aim to provide a convex formulation/relaxation of entropy-constrained compression that directly addresses these limitations. The convex perspective allows us to obtain 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 that gradually constrain the encoder towards a deterministic function and simplify the codebook, we obtain deployable compressors whose empirical entropy closely approaches the convex lower bound and gain a clearer picture of the fundamental limits of our encoders.

Literature and Code Review

The starting point for this work is the classical theory of rate–distortion for discrete memoryless sources, and in particular the Blahut–Arimoto family of algorithms introduced in [1] and [2]. Blahut's 1972 paper [1] and Arimoto's independent work [2] provide iterative procedures for computing channel capacity and rate–distortion functions by updating a conditional distribution $Q(y|x)$ and an output marginal p_Y. These algorithms implicitly exploit the convexity of mutual information in the channel law and can be interpreted as coordinate-descent procedures on a joint convex objective. However, they are typically presented for the standard single-distortion-constraint formulation, without explicit support for mixed-fidelity constraints or a systematic way to attach economic meaning (in bits per unit distortion) to the trade-offs enforced by the constraints.

A complementary influence is the convex-analytic and computational viewpoint emphasized in Peyré and Cuturi's monograph on optimal transport [3]. While their focus is on transport distances rather than mutual information, they provide a modern perspective on entropy-regularized couplings and large-scale convex optimization over probability distributions, which is conceptually close to optimizing over stochastic encoders Q. In this project, rather than reusing an existing codebase, I implement the convex core directly in `CVXPY`, taking advantage of its relative-entropy cone to represent mutual information and of its disciplined convex programming paradigm to enforce global and subset distortion constraints cleanly. The emphasis is not on beating specialized codecs, but on using convex programming as a lens to understand rate–distortion trade-offs, adding additional fidelity constraints, and using convex tool to analyze how R-D is affected by them.

Methods

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

Because the entropy $H(Y)$ of the reconstruction alphabet is nonconvex in Q, we instead minimize the convex surrogate $I(X;Y)$ and later connect it back to $H(Y)$ via deterministic annealing. For a fixed reconstruction alphabet $\mathcal{Y} = \{y_j\}_{j=1}^m$, the core optimization problem is

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

where $\mathcal{S}$ is a fidelity subset with tighter distortion tolerance $D' \ll D$. The objective $I(X;Y)$ is convex in Q (when written via relative entropy), and the feasible set is affine, so the problem is a convex relative-entropy program. The dual variables associated with the global and subset distortion constraints quantify the marginal "price of distortion" in bits per unit distortion and enable a sensitivity analysis of mixed-fidelity trade-offs.

After solving the convex program for a fixed codebook $\mathcal{Y}$, we plan to refine $\mathcal{Y}$ via simple gradient steps $y_j \leftarrow y_j - \eta\nabla_{y_j}\mathbb{E}[d(X, y_j)]$, alternating between convex optimization in Q and codebook updates. To better align the convex information measure $I(X;Y)$ with the entropy $H(Y)$ of a practical (nearly deterministic) encoder, we will introduce a temperature parameter $\tau > 0$ and solve an entropy-regularized version of the problem for a sequence of decreasing temperatures. At high temperature, the encoder remains stochastic; as $\tau \downarrow 0$, the encoder becomes nearly deterministic, $H(Y|X;\tau) \to 0$, and $I(X;Y;\tau) \to H(Y)$. We will additionally prune or merge low-mass reconstruction symbols (small $p_Y(j)$) to further reduce entropy while tracking the gap to the convex lower bound.

All optimization components are implemented in `CVXPY`, which provides convex modeling primitives and solvers for relative-entropy programs. We will evaluate the method both qualitatively and quantitatively: by visualizing rate–distortion (R–D) curves, plotting $I(X;Y)$, $H(Y)$, and $H(Y|X)$ across annealing stages, and examining how dual variables evolve as we vary constraints. Quantitatively, we will report the achieved rate, average distortion, and empirical information quantities, compare against baselines such as entropy-constrained Lloyd quantization, and highlight the benefits of having convex tools that provide certified lower bounds and interpretable dual sensitivities.

Progress Report

So far, I have implemented the basic convex rate–distortion solver with fidelity constraints in `CVXPY` and verified it on several synthetic setups:

- 1D discretized Gaussian and a more challenging 1D trimodal mixture, with squared-error distortion. The resulting R–D curves are decreasing and convex, and the numerical solutions satisfy the constraints and match known Gaussian R–D behavior in the simple case.
- A 2D grid-based Gaussian source with a fixed 2D codebook, again producing sensible R–D curves.
- A mixed-fidelity extension in which a subset S (e.g., points near zero) has a tighter distortion constraint D'. I record dual variables for the global and subset constraints across distortion levels and visualize their sensitivity as a function of achieved distortion.
- Entropy diagnostics that compute $H(Y)$, $H(Y|X)$, and $I_{\text{emp}} = H(Y) - H(Y|X)$ from the solution, confirming that I_{emp} matches the optimization objective I_{opt} up to numerical tolerance.

The current functions live in a python file, and the execution/plots on synthetic data live in a Jupyter notebook, as shown below:

`https://github.com/mnasser3/Convex_lossy_data_compression/tree/main`

once the repository is cleaned up.

For the remaining weeks, I plan to (i) implement a temperature-based deterministic annealing schedule and track $(I(X;Y), H(Y), H(Y|X))$ as temperature decreases, (ii) add basic codebook refinement and pruning strategies, (iii) compare final deterministic encoders against nonconvex baselines while reporting the gap between the convex lower bound and the empirical entropy of the compressor, and (iv) showcase in the report all the advantages of using convex tools and our heuristic convex formulation.

References

References

[1] R. E. Blahut, "Computation of Channel Capacity and Rate–Distortion Functions," *IEEE Transactions on Information Theory*, vol. 18, no. 4, pp. 460–473, 1972.

[2] S. Arimoto, "An Algorithm for Calculating the Capacity of Arbitrary Discrete Memoryless Channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 14–20, 1972.

[3] G. Peyré and M. Cuturi, *Computational Optimal Transport*, Foundations and Trends in Machine Learning, vol. 11, no. 5–6, pp. 355–607, 2019.