Student Project — Maximum Entropy Estimation

I. MAXIMUM ENTROPY ESTIMATION

The problem of estimating an unknown probability density based on the knowledge of some of its moments has important applications in various areas of physics, engineering, and signal processing in particular. One approach toward this estimation is the so-called *maximum entropy approach*[1, p.256 ff.], which formally is concerned with the following problem:
For a compact set $R \subset \mathbb{R}$, let $f \in \mathrm{L}^1(R)$ be a probability density only known via its first n moments $m_0, \ldots, m_{n-1}$, where $m_k := \int_R x^k f(x)\,\mathrm{d}x$. From that partial knowledge one wishes (a) to provide an estimate f_n of f such that the first n moments of f_n match those of f, and (b) analyze the asymptotic behaviour of f_n when $n \to \infty$. An elegant methodology is to search for f_n in a (finitely) parametrized family $\{f_n(\lambda, x)\}$ of functions, and optimize over the unknown parameters λ via a suitable criterion that entails the moments information. For instance, one may wish to select an estimate f_n that maximizes some appropriate cost function, called the *entropy*. Translated into the optimization framework, one is interested in the problem

$$\mathsf{P}(y) : \begin{cases} \sup\limits_{g \in \mathrm{L}^1(R)} & -\int_R g(x) \log g(x)\,\mathrm{d}x \\ \text{s.t.} & \int_R x^j g(x)\,\mathrm{d}x = y_j, \quad j = 0, \ldots, n-1 \\ & g \geq 0, \end{cases} \tag{1}$$

where the vector $y \in \mathbb{R}^n$ represents the given moments. Problem (1) is a convex optimization problem, however the decision variables are infinite-dimensional, which makes the problem computationally intractable. By the duality theory of convex programming and due to the specific problem structure, one can show [1, p. 258] that the problem (1) is equivalent to the convex finite-dimensional optimization problem

$$\mathsf{P}(y) : \sup_{\lambda \in \mathbb{R}^n} \left\{ \langle y, \lambda \rangle - \int_R f_n(\lambda, x)\,\mathrm{d}x \right\}, \qquad \text{where } f_n(\lambda, x) := \exp \sum_{j=0}^{n-1} \lambda_j x^j. \tag{2}$$

It therefore remains to study how to solve (2) efficiently. Since problem (2) is unconstrained one efficient method to solve (2) is the Newton method, which however requires knowledge of the gradient and the Hessian of the objective function in (2). It can be shown [1, Lemma 12.6] that the gradient and the Hessian are given by moments of the density $f_n(\lambda, x)$. As such one aims to study how to compute these moments efficiently. One possible efficient technique for this particular moment computation that has been presented in the literature is Lasserre's semidefinte relaxation hierarchy [1].

II. GOALS OF THE PROJECT

The aim of this project is to investigate potential methods to efficiently compute the above mentioned moments and as such solve the maximum entropy estimation problem. One promising approach would be to study Gauss quadrature methods. In particular we would be interested in exploring the fact that the Gauss quadrature can also be obtained as the solution of a linear program [2] and seeing its performance in terms of computational efficiency. Of course other integration methods, depending on the students interest, could be analyzed.
The mentioned maximum entropy estimation problem is the key ingredient of a recent approximation scheme to channel capacities [3], which, roughly speaking, characterize the amount of

information that can be transmitted reliably over a noisy channel in the information theory literature. We would be curious to see how the method developed in this student project works in the context of approximating channel capacities.

III. REQUIREMENTS

The project is well suited for a student that enjoys mathematics. A solid background in analysis and convex optimization is required.

IV. SUPERVISORS

- Tobias Sutter, Automatic Control Laboratory, ETHZ
- David Sutter, Institute for Theoretical Physics, ETHZ
- Peyman Mohajerin Esfahani, Automatic Control Laboratory, ETHZ
- Prof. John Lygeros, Automatic Control Laboratory, ETHZ

[1] J. B. Lasserre, *Moments, Positive Polynomials and Their Applications*, Imperial College Press Optimization Series, Vol. 1 (Imperial College Press, London, 2010) pp. xxii+361.

[2] Ernest K. Ryu and Stephen P. Boyd, "Extensions of gauss quadrature via linear programming," Foundations of Computational Mathematics , 1–19 (2014).

[3] Tobias Sutter, David Sutter, Peyman Mohajerin Esfahani, and John Lygeros, "Efficient approximation of channel capacities," Information Theory, IEEE Transactions on **61**, 1649–1666 (2015), available at `arXiv:1407.7629`.