Convex optimization and a relative entropy problem

(notes by Michael Baer)

Relative entropy, or Kullback-Leibler divergence, is defined as

$$D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

where the base of log depends on the units in question. It is generally thought of in terms of the penalty of coding according to probability q when the true probability is p. As such, many interesting problems in and properties of information theory include relative entropy.

Consider the following problem: We are made, though ignorance or circumstance, to code according to probability distribution q. At some point we know that the real distribution is within R of probability distribution s; that is, it is some unknown p such that $D(p\|s) \leq R$. What then is the worst-case p? Formally, given positive R and probability mass functions q and s, what is the value of p which achieves

$$\max_{\{p \ | \ D(p\|s) \leq R\}} D(p\|q)?$$

This does not lend itself toward a straightforward Lagrangian solution, and it is not a convex minimization (or concave maximization), so, even though the domain and function are both convex, it is not a convex optimization. Nevertheless, under certain conditions, it can be transformed into one.

Consider those problems in which there exists no candidate $p(x)$ on the border of the simplex — that is, no x and p such that $p(x) = 0$ and $D(p\|s) \leq R$. Then the border of the constraint is fully characterized by $D(p\|s) = R$.

Note that the objective $D(p\|q)$ is convex in p. Therefore, the maximum value is achieved on the border and this problem is equivalent to solving the problem constrained to $D(p\|s) = R$. This, in turn, is equivalent to solving the minimization of $R - D(p\|q) = D(p\|s) - D(p\|q) = \mathbb{E}_{p(x)}[\log(s(x)) - \log(q(x))]$ on this border, or, because this formulation of the problem is linear in p (and thus both convex and concave), over $D(p\|s) \leq R$. In other words,

$$\begin{aligned}
\max_{\{p \ | \ D(p\|s) \leq R\}} D(p\|q) &= \max_{\{p \ | \ D(p\|s) = R\}} D(p\|q) \\
&= R - \min_{\{p \ | \ D(p\|s) = R\}} D(p\|s) - D(p\|q) \\
&= R - \min_{\{p \ | \ D(p\|s) \leq R\}} \mathbb{E}_{p(x)}[\log(s(x)) - \log(q(x))].
\end{aligned}$$

This, then, is a convex optimization problem (minimizing a linear function over a convex domain) and is amenable to methods of solving such problems.

It is worthwhile to note that this analysis applies equally to Euclidean distance, although the resulting linear objective is different and the problem is not as much of a challenge to solve via other means.