Sparse Fitting of Higher-Degree Ising Hamiltonians

Andrew Moore

Summer 2023

Consider an Ising circuit f with N inputs, M outputs, and no auxilliaries. The total spin space is thus $\Sigma := \Sigma^N \times \Sigma^M \cong \mathbb{Z}_2^{N+M}$. Now, we would like to design and optimization problem that searches for the Hamiltonian polynomial on Σ of maximum sparsity and minimum degree, subject to the "hypercube" constraint set

$$\mathcal{H} := \{H(\xi_1) - H(\xi_0) > 0 : \xi_0, \xi_1 \in I_\sigma, L(\xi_1, [\sigma, f(\sigma)]) = 1 + L(\xi_0, [\sigma, f(\sigma)]), \sigma \in \Sigma^M\} \quad (1)$$

(i.e. each correct answer is the only optima of its input level) where L is the Hamming distance function. Because we know that the degree is likely small and we have the computing resources to brute force it, we will ignore optimizing over the degree. Instead, we will fix the degree at d and consider the problem of minimizing the zero-norm of the coefficients t of a degree d Hamiltonian polynomial on Σ, subject to the constraint set $Mt > 0$ where $M = \{v(\xi_1) - v(\xi_0)\}$ is modeled on $\mathcal{H}$ and v is the degree d virtual spin state (note that $H(\xi) = \langle v(\xi), t \rangle$). It is worth noting that since ξ_1 and ξ_0 can only differ by a single bit, the matrix M must be extremely sparse. In fact, many of its columns will be zero, since monomial terms containing only the input spins are constant on each input level. In pratice, we will remove these to make computations faster.

1 L_0 Approximation

The most naive approach to sparse fitting of t is the optimization problem $\min_t \|t\|_0 : Mt \geq 1$. However, since the L_0 norm is not convex, this is in general NP-hard. The easiest approach to approximating the L_0 norm is the linear problem $\min_t \|t\|_1 : Mt \geq 1$, which can be phrased as the traditional linear programming problem $\min_{t,y} \langle 1, y \rangle : Mt \geq 1, t \leq y, t + y \geq 0$. This does work, but produces results which are far from optimal. This section will be devoted to constructing an optimization algorithm which attempts to minimize the L_0 norm directly with a series of smooth approximations by using projected gradient descent.

Let $C := \{x \in \mathbb{R}^n : Mx \geq 1\}$ be the constraint region, and P_C represent orthogonal projection to C (which is well-defined since C is convex). Consider the family of approximate L_0 norms given by the parameterized function[1]

$$f(x; \sigma) := \sum_{i=1}^{n} \frac{x_i^2}{x_i^2 + \sigma} \quad (2)$$

---

[1] c.f. Wei et. al., *Gradient Projection with Approximate L_0 Norm Minimization for Sparse Reconstruction in Compressed Sensing*

As $\sigma \to 0$, $f(x; \sigma) \to \|x\|_0$, but becomes increasingly nonconvex. Our strategy will be gradient descent along f, slowly reducing σ, and projecting back to C after each iteration:

$$z^{(k+1)} := x^{(k)} - \eta \nabla_x f(x^{(k)}; \gamma^k \sigma) \tag{3}$$

$$x^{(k+1)} := P_C(z^{(k+1)}) \tag{4}$$

Where η and γ are hyperparameters defining the step size and the rate of approach to the zero norm respectively. For the sake of making the implementation completely explicit, we calculate the gradient:

$$\frac{\partial}{\partial x_i} f(x; \beta) = \frac{\partial}{\partial x_i} \frac{x_i^2}{x_i^2 + \beta} = \frac{2x_i}{x_i^2 + \beta} + (-1)(x_i^2)(2x_i)(x_i^2 + \beta)^{-2} \tag{5}$$

$$= \frac{2x_i(x_i^2 + \beta) - 2x_i^3}{(x_i^2 + \beta)^2} = \frac{2\beta x_i}{(x_i^2 + \beta)^2} \tag{6}$$

Unfortunately, the function P_C is not necessarily simple, as orthogonal projection to a convex polyhedral cone remains an open problem. It can be expressed as a linearly constrained quadratic programming problem:

$$P_C(x) = \arg\min_y \|x - y\|_2^2 \qquad\qquad My \geq 1 \tag{7}$$

We introduce the Lagrangian $\mathcal{L}(y, \alpha) = \|x - y\|_2^2 - \langle My - 1, \alpha \rangle$. Now we have the saddle-point problem

$$\min_y \max_\alpha \|x - y\|_2^2 - \langle My - 1, \alpha \rangle \qquad\qquad \alpha \geq 0 \tag{8}$$

The dual problem is therefore

$$\max_\alpha \min_y \|x - y\|_2^2 - \langle My - 1, \alpha \rangle \qquad\qquad \alpha \geq 0 \tag{9}$$

Now, we can explicitly evaluate the inner minimization: note that

$$\nabla_y \mathcal{L}(y, \alpha) = \nabla_y (\|x\|_2^2 + \|y\|_2^2 - 2\langle x, y \rangle - \langle y, M^T \alpha \rangle + \langle 1, \alpha \rangle) \tag{10}$$

$$= 2y - 2x - M^T \alpha \implies y^* = x + \frac{1}{2} M^T \alpha \tag{11}$$

$$\implies \min_y \mathcal{L}(y, \alpha) = \mathcal{L}(y^*, \alpha) = \frac{1}{4} \|M^T \alpha\|_2^2 - \left\langle Mx + \frac{1}{2} MM^T \alpha - 1, \alpha \right\rangle \tag{12}$$

$$= \frac{1}{4} \|M^T \alpha\|_2^2 - \langle Mx - 1, \alpha \rangle - \frac{1}{2} \langle M^T \alpha, M^T \alpha \rangle = -\frac{1}{4} \|M^T \alpha\|_2^2 - \langle Mx - 1, \alpha \rangle \tag{13}$$

This yields the final dual problem

$$\min_\alpha \|M^T \alpha\|_2^2 + 4\langle Mx - 1, \alpha \rangle \qquad\qquad \alpha \geq 0 \tag{14}$$

This can now easily be solved by gradient descent with weight clipping, since projection to the non-negative orthant is a simple as $\max(\alpha, 0)$.