Distributionally Robust Optimization with Bias and Variance Reduction
Ronak Mehta, Vincent Roulet, Krishna Pillutla, Zaid Harchaouii

**Algorithm: Prospect**

**Objective:**
\[
\min \max \left\{ \frac{1}{n} \sum_{i=1}^{n} \ell_i(w) \right\}
\]

**Input:**
- \(\eta\): learning rate
- \(w\): stored in memory: current primal iterate \(w_1, \ldots, w_n\)
- \(q\): estimates of loss/gradient at each data point \(q_1, \ldots, q_n\)
- \(\nu\): distribution to project to \(\nu_1, \ldots, \nu_n\)
- \(\sigma\): ambiguity set

**Main Loop:**
1. Sample data point: \(i \sim \text{Unif}\{1, \ldots, n\}\)
2. Compute gradient estimate: \(v_i = q_i \nabla \ell_i(w) + \mu w\)
3. Compute variance reducer: \(v_2 = g_i - \sum_{j=1}^{n} q_j v_j\)
4. Main Update: \(w \leftarrow w - \eta (v_i - v_2)\)
5. Update all tables: \(\tilde{q}_i \leftarrow q_i, \tilde{l}_i \leftarrow \ell_i(w), g_i \leftarrow \nabla \ell_i(w)\)

**Convergence Analysis**

**Assume** the losses are convex, \(L\)-smooth and \(G\)-Lipschitz.

**Define** the condition numbers \(\kappa_f = L / \mu + 1\) and \(\kappa_s = n \sigma_{\max}^2\) and constant \(\kappa_G = G^2(\mu\sigma)\).

1. Prospect with \(\eta \sim \text{poly}(n, \kappa_f, \kappa_s, \kappa_G)\) converges linearly.
2. If in addition, \(\kappa_s \leq 1/6\), then for \(\eta \sim (\kappa_f (L + \mu))^{-1}\) and \(\tau \sim n + \kappa_s\), we have for iterates \(w_0, w_1, \ldots\), that

\[
\mathbb{E}[\|w_n - w^*\|^2_2 \leq 6n^2 \exp(-t/\tau)\|w_0 - w^*\|^2_2]
\]

**Key Idea:** Instead of solving dual problem using true losses (which cost \(n\) oracle calls to compute), use lazily updated table of losses to approximate dual solution. Update direction has terms of gradient evaluations, across datasets.

**Experiments**

- **Training Distributionally Robust Linear Models**
  - Prospect has best/closest to best optimization performance in terms of gradient evaluations, across datasets.
  - Dataset 1
  - Dataset 2
  - Dataset 3

- **Subpopulation Shift in Image/Text Classification**
  - Worst group test error mitigated by Prospect solution on Amazon Reviews.

- **Group Fairness**
  - Median group test error mitigated by Prospect solution on iWildCam.

**Using SRMs to Promote Group Fairness**

- **Full Paper + Code**
  - DR objective correlates with statistical parity fairness score.

**Distributional Empirical Risk Minimization**

- **Model Parameters**
  - \(\beta\): model parameters (primal variables)
  - \(w\): training data weights

- **Distributionally Robust (DR) Objectives**
  - \(\min \max \{\frac{1}{n} \sum_{i=1}^{n} \ell_i(w)\\}\)

- **Convexity**
  - Most “skewed” weights possible

- **Uncertainty Set**
  - Optimal dual variables tend toward vertex

**Goal:** construct a stochastic, linearly convergent optimization algorithm for DR objectives.