POVERVIEW: Convergence Analysis of Sequential Minimal Optimization

Motivation:
- Support vector machines (SVMs) are widely used in many applications.
- Sequential minimal optimization (SMO) has been a popular dual 2-coordinate ascent method for training SVMs for around 20 years.
- SMO can train SVMs with an unregularized bias, which is preferred when there are imbalanced class labels.

New convergence analysis of SMO with uniformly random coordinate selection (rSMO).

This unregularized bias leads to a linear equality constraint across the dual variables,

\[ \sum_{i=1}^{n} \alpha_i x_i = 0 \]  

in addition to the constraints  

\[ 0 \leq x_i \leq c \quad \text{for all } i \in \{1, 2, \ldots, n\} \]

which complicates analysis of modern stochastic dual coordinate ascent (SDCA) methods.

Our argument closely follows the analysis of Necoara & Patrascu.

Sequential Minimal Optimization

- On each iteration, SMO chooses a block from the candidate block set

\[ B \subset \{ \{i\}, \{i, j\} \mid \{i, j\} \in \mathcal{K}, i \neq j \} \]

- The iteration update corresponds to

\[ \begin{align*}
  x^{k+1}_i &= x_i - H^n_i + H^j_j + 2H^j_j \nabla f(x^k + \nabla f(x^k)) \\
  x^{k+1}_j &= x_j - H^n_j + H^j_i + 2H^j_j \nabla f(x^k + \nabla f(x^k))
\end{align*} \]

Block Coordinate Descent

- SMO is an instance of general BCD with an iteration update given by

\[ x^{k+1} = \arg \min_{(y, z) \in \mathcal{X}^2} \left\{ f(y) + \langle \nabla f(z), y - x^k \rangle + \frac{L}{2} \| y - x^k \|_2^2 \right\} \]

- The candidate block set \( B \) contains the supports of the elementary vectors of \( \text{null}(L) \).

Linear Convergence

Our argument closely follows the analysis of Necoara & Patrascu.

- By Lipschitz-continuity of \( \nabla f \) (2) and the property \( H^k \leq L I \) (5),

\[ \mathbb{E}[f(x^{k+1})] \leq f(x^k) + \frac{1}{|B|} \sum_{B \in \mathcal{B}} \min_{d_k \in \mathcal{D}(x^k, B)} \left\{ \langle \nabla f(x^k), d_k \rangle + \frac{L}{2} \| d_k \|_2^2 \right\} . \]

- By \( d_k \) containing the supports of the elementary vectors and properties of conformity,

\[ \mathbb{E}[f(x^{k+1})] \leq f(x^k) + \frac{1}{|B|} \sum_{d_k \in \mathcal{D}(x^k, B)} \left\{ \langle \nabla f(x^k), d_k \rangle + \frac{L}{2} \| d_k \|_2^2 \right\} \]

- By prox-PL inequality (3),

\[ \mathbb{E}[f(x^{k+1})] \leq f(x^k) + \frac{1}{|B|} \sum_{d_k \in \mathcal{D}(x^k, B)} \left\{ \langle \nabla f(x^k), d_k \rangle + \frac{L}{2} \| d_k \|_2^2 \right\} \]

Linear Convergence

We can subtract \( f^* \) from both sides and rearrange the terms, and apply this recursively to get the linear convergence result (6).

Support Vector Identification

- We additionally require that the active set of the solution set \( X^* \) is unique, and the non-degeneracy conditions:

\[ \nabla f(x^*) \neq 0 \quad \text{for all } i \in Z \text{ and } x^* \in X^* \]

\[ \nabla f(x^*) \neq \nabla f(x^*) \quad \text{for all } i \in Z, j = 1, \ldots, n, i \neq j \text{ and } x^* \in X^* \]

- SMO for problems satisfying the additional requirements above detects the final set of support vectors after some finite iterate \( K \).

Intuition:
- Use induction on decreasing order of \( \langle \nabla f(x^*), z \rangle \), for \( i \in Z \) to show that SMO detects the active set after some finite iterate \( K \).

\[ x^0 = \begin{pmatrix} x^0_1 \\ x^0_2 \end{pmatrix} \]

\[ x^k = \begin{pmatrix} x^k_1 \\ x^k_2 \end{pmatrix} \]

\[ x^\infty = \begin{pmatrix} x^\infty_1 \\ x^\infty_2 \end{pmatrix} \]

- Pick out the indices \( i \) that are not on the active set.