Optimization Methods (CS1.404), Spring 2024
Lecture 14

Naresh Manwani

Machine Learning Lab, IIIT-H

March 4th, 2024

Basic Conjugate Direction Algorithm

Given starting point $\mathbf{x}_0$ and H conjugate directions $\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_{n-1} \in \mathbb{R}^n$, the Conjugate Direction Algorithm works as follows:

- For $(k = 0, 1, \ldots, n-1)$
  - $\nabla f(\mathbf{x}_k) = H\mathbf{x}_k + \mathbf{c}$
  - $\alpha_k = -\frac{\nabla f(\mathbf{x}_k)^T \mathbf{d}_k}{\mathbf{d}_k^T H \mathbf{d}_k}$
  - $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$

Theorem

Consider minimization problem $\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2}\mathbf{x}^T H \mathbf{x} + \mathbf{c}^T \mathbf{x}$, where H is symmetric positive definite matrix. For any starting point $\mathbf{x}_0$ and H conjugate directions $\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_{n-1} \in \mathbb{R}^n$, the **Basic Conjugate Direction Algorithm** converges to the unique $\mathbf{x}^*$ (that solves $H\mathbf{x}^* + \mathbf{c} = \mathbf{0}$) in n-steps; that is $\mathbf{x}_n = \mathbf{x}^*$.

Basic Conjugate Algorithm: Example

- $f(x_1, x_2) = 4x_1^2 + x_2^2 - 2x_1x_2$
- **Step 1:** $\mathbf{d}_0 = (1,0)^T$, $\mathbf{x}_0 = (-1,-1)^T$
  - Find $x_1 = x_0 + \alpha_0\mathbf{d}_0$ where $\alpha_0 = \arg\min_{\alpha>0} \; f(\mathbf{x}_0 + \alpha\mathbf{d}_0)$
  - Let $\phi(\alpha) = f(\mathbf{x}_0 + \alpha\mathbf{d}_0) = f(-1+\alpha, -1) = 4(\alpha-1)^2 + 1 + 2(\alpha-1)$
  - $\phi'(\alpha) = 0 \Rightarrow 8(\alpha-1) + 2 = 0 \Rightarrow \alpha_0 = \frac{3}{4}$
  - $\mathbf{x}_1 = \mathbf{x}_0 + \alpha_0\mathbf{d}_0 = (-1,-1)^T + \frac{3}{4}(1,0)^T = (-\frac{1}{4}, -1)^T$.
- **Step 2:** Choosing $\mathbf{d}_1 = (1,4)^T$, as it becomes H-conjugate for $\mathbf{d}_0$.
  - Find $x_2 = x_1 + \alpha_1\mathbf{d}_1$ where $\alpha_1 = \arg\min_{\alpha>0} \; f(\mathbf{x}_1 + \alpha\mathbf{d}_1)$
  - Let $\phi(\alpha) = f(\mathbf{x}_1 + \alpha\mathbf{d}_1) = f(-\frac{1}{4}+\alpha, -1+4\alpha) =$ $4(\alpha-\frac{1}{4})^2 + (4\alpha-1)^2 - 2(\alpha-\frac{1}{4})(4\alpha-1) = \frac{3}{4}(4\alpha-1)^2$
  - $\phi'(\alpha) = 0 \Rightarrow \alpha_1 = \frac{1}{4}$
  - $\mathbf{x}_2 = \mathbf{x}_1 + \alpha_1\mathbf{d}_1 = (-\frac{1}{4}, -1)^T + \frac{1}{4}(1,4)^T = (0,0)^T$.
- Because f is quadratic function in two variables, $\mathbf{x}_2 = \mathbf{x}^*$.

Expanding Subspace Theorem

Theorem

Let $\mathbf{g}_k = \nabla f(\mathbf{x}_k)$. In the Conjugate Direction algorithm,

- $\mathbf{g}_{k+1}^T \mathbf{d}_i = 0$ for all k, $0 \leq k \leq n-1$, and $0 \leq i \leq k$.
- $\mathbf{x}_{k+1} = \arg\min \; f(\mathbf{x})$ s.t. $\mathbf{x} \in \mathbf{x}_0 + \mathcal{B}_k$.

- Let $\mathcal{B}_k$ be the subspace spanned by $\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_k$
- By this lemma, $\mathbf{g}_{k+1}$ is orthogonal to any vector from the subspace spanned $\mathcal{B}_k$.

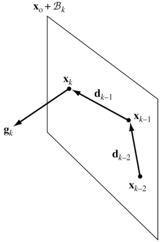

Figure: Illustration of Lemma

The Conjugate Gradient Algorithm

- Conjugate direction method uses pre-specified directions.
- Conjugate gradient algorithm does not use pre-specified directions.
- At each stage, the conjugate gradient algorithm, the direction is calculated as a linear combination of the previous direction and the current gradient in such a way that all the directions are mutually H-conjugate.

The Conjugate Gradient Algorithm

- Consider the quadratic function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T H\mathbf{x} + \mathbf{c}^T\mathbf{x}$, where $\mathbf{x} \in \mathbb{R}^n$ and H is a symmetric positive definite $n \times n$ matrix.
- At $\mathbf{x}_0$, we choose $\mathbf{d}_0$ as the steepest descent direction. I.e., $\mathbf{d}_0 = -\nabla f(\mathbf{x}_0) = -\mathbf{g}_0$.
- Thus, $\mathbf{x}_1 = \mathbf{x}_0 + \alpha_0\mathbf{d}_0$, where $\alpha_0 = \arg\min_{\alpha \geq 0}\ f(\mathbf{x}_0 + \alpha\mathbf{d}_0) = -\frac{\mathbf{g}_0^T\mathbf{d}_0}{\mathbf{d}_0 H\mathbf{d}_0}$.
- Next, we search direction $\mathbf{d}_1$ that is H conjugate of $\mathbf{d}_0$
- We choose $\mathbf{d}_1$ as linear combination of $\mathbf{d}_0$ and $\mathbf{g}_1$.
- In general, at the step $(k+1)$, we choose $\mathbf{d}_{k+1}$ as linear combination of $\mathbf{d}_k$ and $\mathbf{g}_{k+1}$.
- Specifically, we choose $\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_k\mathbf{d}_k,\ k = 0, 1, 2, \ldots$.
- The coefficients $\beta_k,\ k = 0, 1, 2, \ldots$, are chosen such that $\mathbf{d}_{k+1}$ is H-conjugate to $\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_k$.
- This is accomplished by choosing $\beta_k = \frac{\mathbf{g}_{k+1}^T H\mathbf{d}_k}{\mathbf{d}_k^T H\mathbf{d}_k}$.

The Conjugate Gradient Algorithm

For quadratic function, $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T H\mathbf{x} + \mathbf{c}^T\mathbf{x}$, where $\mathbf{x} \in \mathbb{R}^n$ and H is a symmetric positive definite $n \times n$ matrix.

The Conjugate Gradient Algorithm

- **Initialize:** $\mathbf{x}_0$, $\epsilon > 0$, $\mathbf{d}_0 = -\mathbf{g}_0$, $k = 0$
- While $(\|\mathbf{g}_k\| > \epsilon)$
  - Choose $\alpha_k = \arg\min_{\alpha \geq 0} \; f(\mathbf{x}_k + \alpha\mathbf{d}_k) = -\frac{\mathbf{g}_k^T\mathbf{d}_k}{\mathbf{d}_k H\mathbf{d}_k}$
  - $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\mathbf{d}_k$
  - $\mathbf{g}_{k+1} = \nabla f(\mathbf{x}_{k+1}) = H\mathbf{x}_{k+1} + \mathbf{c}$
  - $\beta_k = \frac{\mathbf{g}_{k+1}^T H\mathbf{d}_k}{\mathbf{d}_k^T H\mathbf{d}_k}$
  - $\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_k\mathbf{d}_k$
  - $k = k + 1$
- **Output:** $\mathbf{x}^* = \mathbf{x}_k$, global minimum of $f(\mathbf{x})$.

Conjugate Gradient Algorithm: Conjugate Property of Directions Generated

Proposition

In the conjugate gradient algorithm, the directions $\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_{n-1}$ are H-conjugate.