Conditional Gradient Method, plus Subgradient Optimization

Robert M. Freund

March, 2004

1 The Conditional-Gradient Method for Constrained Optimization (Frank-Wolfe Method)

We now consider the following optimization problem:

$$P: \quad \text{minimize}_x \quad f(x)$$

$$\text{s.t.} \qquad x \in C \ .$$

We assume that $f(x)$ is a convex function, and that C is a convex set. Herein we describe the conditional-gradient method for solving P, also called the Frank-Wolfe method. This method is one of the cornerstones of optimization, and was one of the first successful algorithms used to solve nonlinear optimization problems. It is based on the premise that the set C is well-suited for linear optimization. That means that either C is itself a system of linear inequalities $C = \{x \mid Ax \leq b\}$, or more generally that the problem:

$$LO_c: \quad \text{minimize}_x \quad c^T x$$

$$\text{s.t.} \qquad x \in C$$

is easy to solve for any given objective function vector c.

This being the case, suppose that we have a given iterate value $\bar{x} \in C$. Let us linearize the function $f(x)$ at $x = \bar{x}$. This linearization is:

$$z_1(x) := f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) \ ,$$

which is the first-order Taylor expansion of $f(\cdot)$ at $\bar{x}$. Since we can easily do linear optimization on C, let us solve:

$$LP: \quad \text{minimize}_x \quad z_1(x) = f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x})$$

$$\text{s.t.} \qquad x \in C \ ,$$

which simplifies to:

$$LP: \quad \text{minimize}_x \quad \nabla f(\bar{x})^T x$$

$$\text{s.t.} \qquad x \in C \ .$$

Let x^* denote the optimal solution to this problem. Then since C is a convex set, the line segment joining $\bar{x}$ and x^* is also in C, and we can perform a line-search of $f(x)$ over this segment. That is, we solve:

$$LS: \quad \text{minimize}_\alpha \quad f(\bar{x} + \alpha(x^* - \bar{x}))$$

$$\text{s.t.} \qquad 0 \leq \alpha \leq 1 \ .$$

Let $\bar{\alpha}$ denote the solution to this line-search problem. We re-set $\bar{x}$:

$$\bar{x} \leftarrow \bar{x} + \bar{\alpha}(x^* - \bar{x})$$

and repeat this process.

The formal description of this method, called the conditional gradient method (or the Frank-Wolfe) method, is:

Step 0: Initialization. Start with a feasible solution $x^0 \in C$. Set $k = 0$. Set $LB \leftarrow -\infty$.

Step 1: Update upper bound. Set $UB \leftarrow f(x^k)$. Set $\bar{x} \leftarrow x^k$.

Step 2: Compute next iterate.

- Solve the problem

$$\begin{array}{rcll} \bar{z} & = & \min_x & f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) \\ & & \text{s.t.} & x \in C \ , \end{array}$$

  and let x^* denote the solution.
- Solve the line-search problem:

$$\begin{array}{ll} \text{minimize}_\alpha & f(\bar{x} + \alpha(x^* - \bar{x})) \\ \text{s.t.} & 0 \le \alpha \le 1 \ , \end{array}$$

  and let $\bar{\alpha}$ denote the solution.
- Set $x^{k+1} \leftarrow \bar{x} + \bar{\alpha}(x^* - \bar{x})$

Step 3: Update Lower Bound. Set $LB \leftarrow \max\{LB, \bar{z}\}$.

Step 4: Check Stopping Criteria. If $|UB - LB| \le \epsilon$, stop. Otherwise, set $k \leftarrow k + 1$ and go to **Step 1**.

The upper bound values UB are simply the objective function values of the iterates $f(x^k)$ for $k = 0, \ldots$. This is a monotonically decreasing sequence because the line-search guarantees that each iterate is an improvement over the previous iterate.

The lower bound values LB result from the convexity of $f(x)$ and the gradient inequality for convex functions:

$$f(x) \ge f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) \quad \text{for any } x \in C \ .$$

Therefore

$$\min_{x\in C} f(x) \geq \min_{x\in C} f(\bar{x}) + \nabla f(\bar{x})^T(x-\bar{x}) \ = \bar{z} \ ,$$

and so the optimal objective function value of P is bounded below by $\bar{z}$.

The following theorem concerns convergence of the conditional gradient method:

Theorem 1.1 Conditional Gradient Convergence Theorem *Suppose that C is a bounded set, and that there exists a constant L for which*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|$$

for all $x, y \in C$. Then there exists a constant $\Omega > 0$ for which the following is true:

$$f(x^k) - \min_{x\in C} f(x) \leq \frac{\Omega}{k} \ \textbf{.q.e.d.}$$

1.1 Proof of Theorem 1.1

1.2 Illustration of the Conditional Gradient Method

Consider the following instance of P:

$$\begin{array}{lll} P: & \text{minimize} & f(x) \\ & & \\ & \text{s.t.} & x \in C \ , \end{array}$$

where

$$f(x) = f(x_1, x_2) = -32x_1 + x_1^4 - 8x_2 + x_2^2$$

and

$$C = \{(x_1, x_2) \mid x_1 - x_2 \leq 1, \ 2.2x_1 + x_2 \leq 7, \ x_1 \geq 0, \ x_2 \geq 0\} \ .$$

Notice that the gradient of $f(x_1, x_2)$ is given by the formula:

$$\nabla f(x_1, x_2) = \left(\begin{array}{c} 4x_1^3 - 32 \\ 2x_2 - 8 \end{array} \right) \ .$$

Suppose that $x^k = \bar{x} = (0.5, 3.0)$ is the current iterate of the Frank-Wolfe method, and the current lower bound is $LB = -100.0$. We compute $f(\bar{x}) = f(0.5, 3.0) = -30.9375$ and we compute the gradient of $f(x)$ at $\bar{x}$:

$$\nabla f(0.5, 3.0) = \begin{pmatrix} 4x_1^3 - 32 \\ 2x_2 - 8 \end{pmatrix} = \begin{pmatrix} -31.5 \\ -2.0 \end{pmatrix} .$$

We then create and solve the following linear optimization problem:

$$\begin{array}{lll} LP: & \bar{z} = \min_{x_1,x_2} & -30.9375 - 31.5(x_1 - 0.5) - 2.0(x_2 - 3.0) \\ & & \\ & \text{s.t.} & x_1 - x_2 \leq 1 \\ & & 2.2x_1 + x_2 \leq 7 \\ & & x_1 \geq 0 \\ & & x_2 \geq 0 . \end{array}$$

The optimal solution of this problem is:

$$x^* = (x_1^*, x_2^*) = (2.5, 1.5) ,$$

and the optimal objective function value is:

$$\bar{z} = -50.6875 .$$

Now we perform a line-search of the 1-dimensional function

$$\begin{array}{rcl} f(\bar{x} + \alpha(x^* - \bar{x})) & = & -32(\bar{x}_1 + \alpha(x_1^* - \bar{x}_1)) + (\bar{x}_1 + \alpha(x_1^* - \bar{x}_1))^4 \\ & & -8(\bar{x}_2 + \alpha(x_2^* - \bar{x}_2)) + (\bar{x}_2 + \alpha(x_2^* - \bar{x}_2))^2 \end{array}$$

over $\alpha \in [0, 1]$. This function attains its minimum at $\bar{\alpha} = 0.7165$ and we therefore update as follows:

$$x^{k+1} \leftarrow \bar{x} + \bar{\alpha}(x^* - \bar{x}) = (0.5, 3.0) + 0.7165((2.5, 1.5) - (0.5, 3.0)) = (1.9329, 1.9253)$$

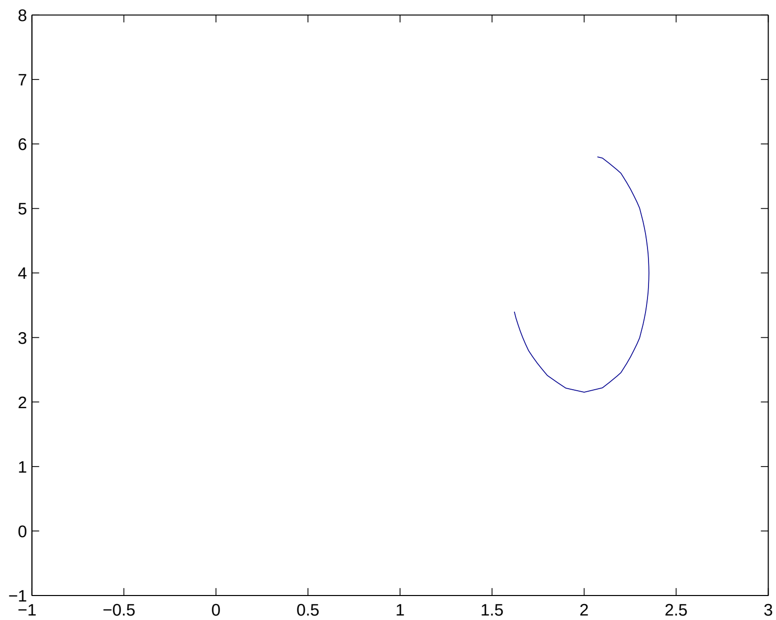
and

$$LB \leftarrow \max\{LB, \bar{z}\} = \max\{-100, -50.6875\} = -50.6875 .$$

The new upper bound is

$$UB = f(x^{k+1}) = f(1.9329, 1.9253) = -59.5901 .$$

This is illustrated in Figure 1.



8
7
6
5
4
3
2
1
0
−1
−1
−0.5
0
0.5
1
1.5
2
2.5
3