The Frank–Wolfe algorithm

Consider the problem to

$$\text{minimize } f(x), \tag{1a}$$
$$\text{subject to } x \in S, \tag{1b}$$

where $S \subset \Re^n$ is a polyhedron and where $f : \Re^n \mapsto \Re$ is a continuously differentiable function. (We suppose that S is also bounded, for simplicity of the presentation only.)

The Frank–Wolfe algorithm works as follows:

0. Choose an *initial solution*, $x_0 \in S$. Let $k := 0$.

   Here one typically chooses an arbitrary basic feasible solution, that is, an extreme point.

1. Determine a *search direction*, p_k.

   In the Frank–Wolfe algorithm one determines p_k through the solution of the approximation of the problem (1) that is obtained by replacing the function f with its first-order Taylor expansion around x_k: therefore, solve the problem to

   $$\text{minimize } z_k(y) := f(x_k) + \nabla f(x_k)^{\mathrm{T}}(y - x_k), \tag{2}$$
   $$\text{subject to } y \in S. \tag{3}$$

   This is an LP problem, and it gives an extreme point, y_k, as an optimal solution. The search direction is $p_k := y_k - x_k$, that is, the direction vector from the feasible point x_k towards the extreme point. Observe that this is a feasible direction, since both x_k and y_k belong to S and S is convex.

2. Determine a *step length*, α_k, such that

   $$f(x_k + \alpha_k p_k) < f(x_k). \tag{4}$$

   Here, we must limit the step length to be at most 1, because for $\alpha > 1$ the solution becomes infeasible; the line search therefore has the form

   $$\underset{\alpha \in [0,1]}{\text{minimize}} \; f(x_k + \alpha p_k).$$

3. New iteration point:
$$x_{k+1} = x_k + \alpha_k p_k.$$
4. If a *stopping criterion* is fulfilled $\longrightarrow$ Stop! x_{k+1} is an approximation of x_*. Otherwise, let $k := k + 1$, and go to 1.

What is left is (a) to motivate why the search direction p_k is a descent direction, so that we are guaranteed to be able to fulfill (4), and (b) to find good termination criteria.

We say that $x \in S$ is stationary if $\nabla f(x)^{\mathrm{T}}(y - x) \geq 0$ for all $y \in X$. This is a necessary condition for x to be a local minimum. Because suppose otherwise, that $x \in S$ is a local minimum for which $\nabla f(x)^{\mathrm{T}}(\bar{y} - x) < 0$ holds for some $\bar{y} \in S$. But $p := \bar{y} - x$ is a feasible direction, so we therefore have that p is a feasible descent direction. This contradicts the local optimality of x.

We can now note that
$$z_k(y_k) \leq z_k(x_k) = 0.$$
(The inequality stems from the fact that y_k solves the LP problem, but that may not be true for x_k.) If $z_k(y_k) = 0$ it means that
$$\nabla f(x_k)^{\mathrm{T}}(y - x_k) \geq 0, \qquad \forall y \in S,$$
that is, x_k is a stationary point. If this is true we terminate the algorithm with a stationary point at hand. Note: if f is convex, then x_k is a global minimum.

If, on the other hand, $z_k(y_k) < 0$ holds, then this means that $\nabla f(x_k)^{\mathrm{T}}(y_k - x_k) < 0$, that is, the vector $p_k := y_k - x_k$ towards the LP solution is a descent direction.

A more valuable termination criterion in practice is given by upper and lower bounds on the optimal objective value. This is valid for convex problems only! That is, what follows is valid if f is a convex function, but not otherwise. The basis for the arguments to follow is the known property that f is convex on S if and only if
$$f(y) \geq f(x) + \nabla f(x)^{\mathrm{T}}(y - x), \qquad x, y \in S.$$
We therefore have that
$$\begin{aligned} z_k(y_k) &= f(x_k) + \nabla f(x_k)^{\mathrm{T}}(y_k - x_k) \\ &\leq f(x_k) + \nabla f(x_k)^{\mathrm{T}}(x_* - x_k) \\ &\leq f(x_*), \end{aligned}$$
where the equality follows from the definition of the optimal solution to the LP problem, the first inequality by the fact that y_k solves this LP problem but not necessarily x_*, and the second inequality from the convexity of f. So, we have that at every iteration k,
$$z_k(y_k) \leq f(x_*) \leq f(x_k),$$

where the last inequality holds since x_k is a feasible solution. In the Frank–Wolfe algorithm the value of f descends after each iteration, that is, the sequence $\{f(x_k)\}$ strictly monotonically decreases towards $f(x_*)$, while the sequence $\{z_k(y_k)\}$ approaches $f(x_*)$ from below—but not necessarily monotonically. We therefore always have an interval, $[z_k(y_k), f(x_k)]$, which contains the optimal objective value. Through this knowledge, we can construct a termination criterion of the type: Let $\varepsilon > 0$ be an a priori chosen value of an acceptable relative objective error. Then, "if $(f(x_k) - z_k(y_k))/|z_k(y_k)| \leq \varepsilon$, then terminate."