WIKIPEDIA

Wolfe conditions

In the unconstrained minimization problem, the **Wolfe conditions** are a set of inequalities for performing **inexact** line search, especially in quasi-Newton methods, first published by Philip Wolfe in 1969.[1][2]

In these methods the idea is to find

$$\min_x f(\mathbf{x})$$

for some smooth $f : \mathbb{R}^n \to \mathbb{R}$. Each step often involves approximately solving the subproblem

$$\min_\alpha f(\mathbf{x}_k + \alpha \mathbf{p}_k)$$

where $\mathbf{x}_k$ is the current best guess, $\mathbf{p}_k \in \mathbb{R}^n$ is a search direction, and $\alpha \in \mathbb{R}$ is the step length.

The inexact line searches provide an efficient way of computing an acceptable step length α that reduces the objective function 'sufficiently', rather than minimizing the objective function over $\alpha \in \mathbb{R}^+$ exactly. A line search algorithm can use Wolfe conditions as a requirement for any guessed α, before finding a new search direction $\mathbf{p}_k$.

Contents

Armijo rule and curvature

A step length α_k is said to satisfy the *Wolfe conditions*, restricted to the direction $\mathbf{p}_k$, if the following two inequalities hold:

$$\textbf{i)} \quad f(\mathbf{x}_k + \alpha_k \mathbf{p}_k) \leq f(\mathbf{x}_k) + c_1 \alpha_k \mathbf{p}_k^{\mathrm{T}} \nabla f(\mathbf{x}_k),$$

$$\textbf{ii)} \quad -\mathbf{p}_k^{\mathrm{T}} \nabla f(\mathbf{x}_k + \alpha_k \mathbf{p}_k) \leq -c_2 \mathbf{p}_k^{\mathrm{T}} \nabla f(\mathbf{x}_k),$$

with $0 < c_1 < c_2 < 1$. (In examining condition (ii), recall that to ensure that $\mathbf{p}_k$ is a descent direction, we have $\mathbf{p}_k^{\mathrm{T}} \nabla f(\mathbf{x}_k) < 0$, as in the case of gradient descent, where $\mathbf{p}_k = -\nabla f(\mathbf{x}_k)$, or Newton–Raphson, where $\mathbf{p}_k = -\mathbf{H}^{-1} \nabla f(\mathbf{x}_k)$ with $\mathbf{H}$ positive definite.)

c_1 is usually chosen to be quite small while c_2 is much larger; Nocedal and Wright[3] give example values of $c_1 = 10^{-4}$ and $c_2 = 0.9$ for Newton or quasi-Newton methods and $c_2 = 0.1$ for the nonlinear conjugate gradient method. Inequality i) is known as the **Armijo rule**[4] and ii) as the **curvature condition**; i) ensures that the step length α_k decreases f 'sufficiently', and ii) ensures that the slope has been reduced sufficiently. Conditions i) and ii) can be interpreted as respectively providing an upper and lower bound on the admissible step length values.

Strong Wolfe condition on curvature

Denote a univariate function φ restricted to the direction $\mathbf{p}_k$ as $\varphi(\alpha) = f(\mathbf{x}_k + \alpha \mathbf{p}_k)$. The Wolfe conditions can result in a value for the step length that is not close to a minimizer of φ. If we modify the curvature condition to the following,

$$\textbf{iii)} \quad \left|\mathbf{p}_k^{\mathrm{T}} \nabla f(\mathbf{x}_k + \alpha_k \mathbf{p}_k)\right| \leq c_2 \left|\mathbf{p}_k^{\mathrm{T}} \nabla f(\mathbf{x}_k)\right|$$

then i) and iii) together form the so-called **strong Wolfe conditions**, and force α_k to lie close to a critical point of φ.

Rationale

The principal reason for imposing the Wolfe conditions in an optimization algorithm where $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha \mathbf{p}_k$ is to ensure convergence of the gradient to zero. In particular, if the cosine of the angle between $\mathbf{p}_k$ and the gradient,

$$\cos\theta_k = \frac{\nabla f(\mathbf{x}_k)^{\mathrm{T}} \mathbf{p}_k}{\|\nabla f(\mathbf{x}_k)\| \|\mathbf{p}_k\|}$$

is bounded away from zero and the i) and ii) conditions hold, then $\nabla f(\mathbf{x}_k) \to 0$.

An additional motivation, in the case of a quasi-Newton method, is that if $\mathbf{p}_k = -B_k^{-1} \nabla f(\mathbf{x}_k)$, where the matrix B_k is updated by the BFGS or DFP formula, then if B_k is positive definite ii) implies B_{k+1} is also positive definite.

References

1. Wolfe, P. (1969). "Convergence Conditions for Ascent Methods". *SIAM Review*. **11** (2): 226–000. doi:10.1137/1011036 (https://doi.org/10.1137%2F1011036). JSTOR 2028111 (https://www.jstor.org/stable/2028111).
2. Wolfe, P. (1971). "Convergence Conditions for Ascent Methods. II: Some Corrections". *SIAM Review*. **13** (2): 185–000. doi:10.1137/1013035 (https://doi.org/10.1137%2F1013035).
3. Nocedal, Jorge; Wright, Stephen (1999). *Numerical Optimization* (https://books.google.com/books?id=epc5fX0lqRIC&lpg=PP1&pg=PA38#v=onepage&q).
4. Armijo, Larry (1966). "Minimization of functions having Lipschitz continuous first partial derivatives" (http://projecteuclid.org/euclid.pjm/1102995080). *Pacific J. Math.* **16** (1): 1–3. doi:10.2140/pjm.1966.16.1 (https://doi.org/10.2140%2Fpjm.1966.16.1).

Further reading