1 Conjugate Gradients, Introduction to The Calculus of Variations (July 21)

1.1 Conjugate Gradient Method

Assume all of the conditions of the previous class.

We will describe a new optimization algorithm that is a type of conjugate direction method. Start at $x_0 \in \mathbb{R}^n$. Choose $d_0 = -g_0 = -\nabla f(x_0) = b - Qx_0$. Recursively define $d_{k+1} = -g_{k+1} + \beta_k d_k$, where $g_{k+1} = Qx_{k+1} - b$ and

$$\beta_k = \frac{g_{k+1}^T Q d_k}{d_k^T Q d_k}$$

and

$$x_{k+1} = x_k + \alpha_k d_k,$$

where

$$\alpha_k = -\frac{g_k^T d_k}{d_k^T Q d_k}.$$

Given an initial point x_0, take $d_0 = -g_0 = b - Qx_0$. By definition, $x_1 = x_0 + \alpha_0 d_0$; we need to find α_0. This is

$$\alpha_0 = -\frac{g_0^T d_0}{g_0^T Q g_0}.$$

Then $x_2 = x_1 + \alpha_1 d_1$. By definition, $\alpha_1 = -\frac{g_1^T d_1}{d_1^T Q d_1}$, where $d_1 = -g_1 + \beta_0 d_0$, where $\beta_0 = \frac{g_1^T Q d_0}{d_0^T Q d_0}$.

Some remarks:

1. Like the other conjugate direction methods, this method converges to the minimizer x_* in n steps.

2. We have a procedure to find the direction vectors d_k.

3. This method makes good *uniform* progress towards the solution at every step.

1.2 Bounds on Convergence

As before, consider $q(x) = \frac{1}{2}(x - x_*)^T Q(x - x_*) = f(x) + \text{const}$. It's better to look at q rather than f, since q behaves like a distance function relative to x_*. (More on this in HW7.)

Theorem 1.1.

$$q(x_{k+1}) \leq \left(\max_{\substack{\lambda \\ \textit{eigval of } Q}} (1 + \lambda P_k(\lambda))^2 \right) q(x_k),$$

where P_k is any polynomial of degree k.

Proof. In the textbook; will not be proven in class. $\square$

For example, suppose Q has $m \leq n$ distinct eigenvalues. Choose a polynomial P_{m-1} such that $1 + \lambda P_{m-1}(\lambda)$ has its m zeroes at the m eigenvalues of Q. With such a polynomial, we would get $q(x_m) \leq 0$, implying that $q(x_m) = 0$; the conjugate gradient method terminates at the mth step, i.e. $x_m = x_*$.

1.3 Introducing The Calculus of Variations

Consider the problem

$$\text{minimize } F[u] \\ u \in \mathcal{A},$$

where $\mathcal{A}$ is a set of functions. Here, F is a function of functions, often called a *functional*. This is the general unconstrained calculus of variations problem.

For example, consider

$$\mathcal{A} = \{u \in C^1([0,1], \mathbb{R}) : u(0) = u(1) = 1\}.$$

Define $F : \mathcal{A} \to \mathbb{R}$ by

$$F[u(\cdot)] \coloneqq \frac{1}{2} \int_0^1 (u(x)^2 + u'(x)^2)\, dx.$$

To solve the minimization problem

$$\text{minimize } F[u] \\ u \in \mathcal{A}$$

is to find a $u^* \in \mathcal{A}$ such that $F[u^*] \leq F[u]$ for all $u \in \mathcal{A}$. To do so, we will

1. We will derive first order necessary conditions for a minimizer.
2. We will find a function satisfying these conditions.
3. We will check that this function is indeed a minimizer. (This is not always possible.)

(Consider the obvious parallels with finite dimensional optimization.)

Fix $v \in C^1([0,1], \mathbb{R})$ with $v(0) = v(1) = 0$. Suppose u^* is a minimizer of F on $\mathcal{A}$. Clearly $u^* + sv \in \mathcal{A}$ for all $s \in \mathbb{R}$. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(s) \coloneqq F[u^* + sv]$. Then $f(s) \geq f(0)$ for all s, since u^* is a minimizer of F. Then 0 is a minimizer of f, implying $f'(0) = 0$. How do we actually compute $f'(0)$? Since

$$\begin{aligned} f(s) &= \frac{1}{2} \int_0^1 (u^*(x) + sv(x))^2 + (u^{*\prime}(x) + sv'(x))^2\, dx \\ &= \frac{1}{2} \int_0^1 (u^*(x)^2 + u^{*\prime}(x)^2)\, dx + s \int_0^1 (u^*(x)v(x) + u^{*\prime}(x)v'(x))\, dx + \frac{s^2}{2} \int_0^1 (v(x)^2 + v'(x)^2)\, dx, \end{aligned}$$

implying that

$$f'(s) = \int_0^1 (u^*(x)v(x) + u^{*'}(x)v'(x))\,dx + s\int_0^1 (v(x)^2 + v'(x)^2)\,dx,$$

or

$$0 = \int_0^1 (u^*(x)v(x) + u^{*'}(x)v'(x))\,dx \text{ for all } v \in C^1([0,1],\mathbb{R}) \text{ such that } v(0) = v(1) = 0. \qquad (*)$$

The above equality holds for all $v \in C^1([0,1],\mathbb{R})$ such that $v(0) = v(1) = 0$. This is a *primitive* form of the first order necessary conditions.

Let us call the functions v described in (*) the *test functions on* $[0,1]$. We would like to write (*) in a more useful way. Let us make the simplifying assumption that u^* is C^2. Integration by parts gives

$$\int_0^1 u^{*'}(x)v'(x)\,dx = \cancel{u^{*'}(x)v(x)\Big|_0^1} - \int_0^1 u^{*''}(x)v(x)\,dx = \int_0^1 u^{*''}(x)v(x)\,dx.$$

By substituting this into (*) we obtain

$$\int_0^1 (u^*(x)v(x) - u^{*''}(x)v(x))\,dx = 0.$$

Factor the common v out to get

$$\int_0^1 (u^*(x) - u^{*''}(x))v(x)\,dx = 0 \text{ for all test functions } v \text{ on } [0,1].$$

So we have a continuous function $u^*(x) - u^{*''}(x)$ that is zero whenever "integrated against test functions". We claim that any function satisfying this condition must be zero. This result or its variations is called the *fundamental lemma of the calculus of variations*. We shall show that $u^* = u^{*''}$ on $[0,1]$; this gives us the first order necessary conditions we wanted in the first place.

Theorem 1.2. *(Fundamental lemma of the calculus of variations) Suppose $g \in C^0([a,b])$. If*

$$\int_a^b g(x)v(x)\,dx = 0$$

for all test functions v on $[a,b]$, then $g \equiv 0$ on $[a,b]$.

So the first order necessary condition we derived are that $u^* = u^{*''}$ on $[0,1]$, as well as $u^*(0) = u^*(1) = 1$. By MAT267, $u^*(x) = c_1e^x + c_2e^{-x}$ for some constants $c_1, c_2 \in \mathbb{R}$. Some work gives $c_1 = \frac{1}{e+1}$ and $c_2 = \frac{e}{e+1}$. Therefore

$$u^*(x) = \frac{1}{e+1}e^x + \frac{e}{e+1}e^{-x}$$

is the only C^1 minimizer candidate.

We must finally check that u^* is indeed a minimizer. Some work shows that $F[u^* + sv] \geq F[u^*]$ for all $s \in \mathbb{R}$ and all test functions v on $[0,1]$. Choose $u \in \mathcal{A}$. Let $v = u - u^*$; this is a test function on $[0,1]$, so $F[u] \geq F[u^*]$, showing that u^* is, in fact, the global minimizer.