Lecture Outline

- Unimodal functions
- Golden section search
- Fibonacci search

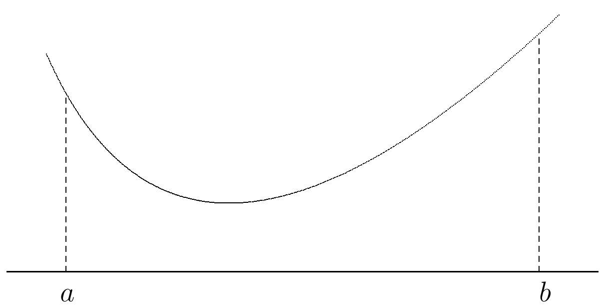
Unimodal functions

A function $f : \Re \to \Re$ is unimodal on $[a, b]$ if

- it has a unique minimizer c over $[a, b]$;
- f is strictly decreasing on $[a, c]$ and strictly increasing on $[c, b]$.

The line search methods are designed to find the minimizer of a unimodal function over a closed interval.

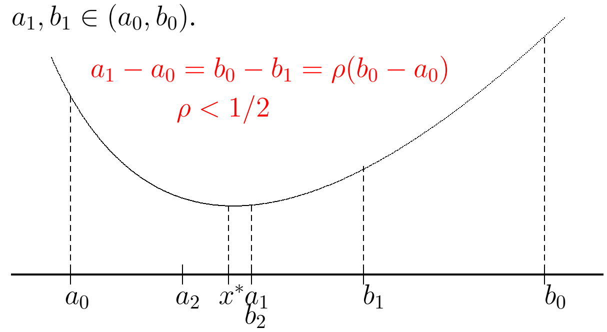
Unimodal functions



A unimodal function

Line search methods

Reduce the search space by locating a smaller interval containing the minimizer: Evaluate f at two points $a_1, b_1 \in (a_0, b_0)$.



Golden section search

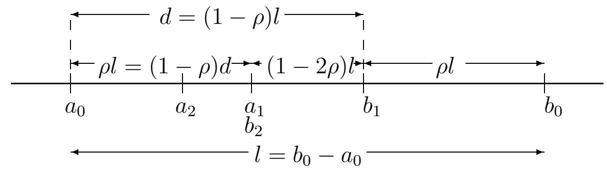
- If $f(a_1) < f(b_1)$, then the minimizer is in the interval $[a_0, b_1]$.
- If $f(a_1) > f(b_1)$ then the minimizer is in $[a_1, b_0]$

$$\Downarrow$$

The range of uncertainty will be reduced by the factor of $(1-\rho)$, and we can continue the search using the same method over a smaller interval.

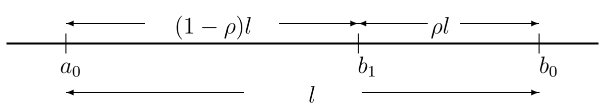
Golden section search

Finding ρ in golden section search



$$\left.\begin{array}{l} d = (1-\rho)l, \quad \Rightarrow \; l = \frac{d}{1-\rho} \\ \text{length of } [a_0, a_1] = [a_0, b_2]: \\ \qquad\qquad \rho l = (1-\rho)d \end{array}\right\} \frac{\rho}{1-\rho} = 1-\rho$$

Golden section search



$$\frac{\rho}{1-\rho} = 1-\rho$$

The ratio of the shorter segment to the longer equals to the ratio of the longer to the sum of the two.

In Ancient Greece, this division was referred to as the *Golden Section*

Golden section search

We can now compute ρ by solving the quadratic equation

$$\rho^2 - 3\rho + 1 = 0.$$

We are looking for $\rho < 1/2$, so the solution is

$$\rho = \frac{3-\sqrt{5}}{2} \approx 0.382.$$

The uncertainty interval is reduced by the fraction of $1-\rho \approx 0.618$ at each step. So, the reduction factor after N steps is

$$(1-\rho)^n \approx 0.618^N.$$

Fibonacci search

- Instead of using the same value of ρ at each step, we can vary it, using a different value ρ_k for each step k.
- Again, we select $\rho_k \in [0, 1/2]$ in the way that only one new function evaluation is required at each step.
- Using reasonings similar to those for the choice of ρ in golden section search, we obtain the following relations for the values of ρ_k:

$$\rho_{k+1} = 1 - \frac{\rho_k}{1-\rho_k}, \quad k = 1, \ldots, n-1.$$

Fibonacci search

- In order to minimize the interval of uncertainty after N steps, we consider the following minimization problem:

$$\begin{array}{ll} \text{minimize} & (1-\rho_1)(1-\rho_2)\cdots(1-\rho_N) \\ \text{subject to} & \rho_{k+1} = 1 - \frac{\rho_k}{1-\rho_k},\ k = 1, \ldots, N-1 \\ & 0 \le \rho_k \le \frac{1}{2},\ k = 1, \ldots, N. \end{array}$$

- The **Fibonacci sequence** $\{F_k, k \ge 0\}$ is defined by $F_0 = F_1 = 1$ and the recursive relation $F_{k+1} = F_k + F_{k-1}$.

Fibonacci search

Theorem 1. *The optimal solution to the above problem is given by*

$$\rho_k = 1 - \frac{F_{N-k+1}}{F_{N-k+2}},\ k = 1, \ldots, N,$$

where F_k is the k^{th} element of the Fibonacci sequence.

Fibonacci search

- We can recursively express all variables $\rho_k,\ k = 1, \ldots, N$ in the objective function through one of the variables, say ρ_N.
- If we denote the resulting univariate function by $f_N(\rho_N)$, then

$$f_N(\rho_N) = \frac{1-\rho_N}{F_N - F_{N-2}\rho_N},\ N \ge 2.$$

We will prove this using induction by N.

Fibonacci search

- For $N = 2$, we have $\rho_1 = \frac{1-\rho_2}{2-\rho_2}$, so

$$f_2(\rho_2) = \left(1 - \frac{1-\rho_2}{2-\rho_2}\right)(1-\rho_2) = \frac{1-\rho_2}{2-\rho_2} = \frac{1-\rho_2}{F_2 - F_0\rho_2}.$$

- Assuming that the statement is correct for some $N = K-1$, *i.e.*,

$$f_{K-1}(\rho_{K-1}) = \frac{1-\rho_{K-1}}{F_{K-1} - F_{K-3}\rho_{K-1}},$$

we need to show that it is also correct for $N = K$.

Fibonacci search

We have $\rho_{K-1} = \frac{1-\rho_K}{2-\rho_K}$, so

$$\begin{aligned} f_K(\rho_K) &= f_{K-1}\left(\frac{1-\rho_K}{2-\rho_K}\right)(1-\rho_K) \\ &= \frac{1 - \frac{1-\rho_K}{2-\rho_K}}{F_{K-1} - F_{K-3}\frac{1-\rho_K}{2-\rho_K}}(1-\rho_K) \\ &= \frac{1-\rho_K}{2F_{K-1} - F_{K-3} - (F_{K-1} - F_{K-3})\rho_K} \\ &= \frac{1-\rho_K}{F_K - F_{K-2}\rho_K}. \qquad \square \end{aligned}$$

Fibonacci search

Next, we will show that $f_N(\rho_N)$ is a strictly decreasing function on $[0, \frac{1}{2}]$. We can do so by showing that the derivative $f'_N(\rho_N) < 0, \forall \rho_N \in [0, \frac{1}{2}]$.
Indeed,

$$f'_N(\rho_N) = \frac{-F_N + F_{N-2}}{(F_N - F_{N-2}\rho_N)^2} = \frac{-F_{N-1}}{(F_N - F_{N-2}\rho_N)^2} < 0, \; \forall \rho_N \le \tfrac{1}{2}.$$

Therefore,

$$\min_{\rho_N \in [0,1/2]} f_N(\rho_N) = f_N(1/2) = \frac{1-1/2}{F_N - F_{N-2}/2} = \frac{1}{F_{N+1}} -$$

the reduction factor after N steps of the Fibonacci search

Fibonacci search

Returning to the original problem, we have

$$\begin{aligned} \rho_N &= 1/2 = 1 - \frac{F_1}{F_2}; \\ \rho_{N-1} &= \frac{1-\rho_N}{2-\rho_N} = \frac{F_1}{F_3} = 1 - \frac{F_2}{F_3}; \\ &\vdots \\ \rho_{k+1} &= 1 - \frac{F_{N-k}}{F_{N-k+1}}; \\ \rho_k &= \frac{1-\rho_{k+1}}{2-\rho_{k+1}} = \frac{F_{N-k}}{F_{N-k+2}} = 1 - \frac{F_{N-k+1}}{F_{N-k+2}}; \\ &\vdots \\ \rho_1 &= 1 - \frac{F_N}{F_{N+1}} \qquad \square \end{aligned}$$

Gradient Methods

Minimize $f(x)$, $x\in\mathbb{R}^n$ $\qquad f(x)\in C^1(\mathbb{R}^n)$

Gradients, level sets ...

$f:\mathbb{R}^n\to\mathbb{R}$. The level set of f corresponding to the level c is given by $S=\{x\in\mathbb{R}^n: f(x)=c\}$

A curve γ in set S is

$$\gamma=\{x(t): t\in(a,b)\}\subset S,$$

where $x(t)$ is a continuous function.

Let $g(t)=f(x(t))$, $t\in(a,b)$

$t_0\in(a,b)$; $x_0=x(t_0)$

Since $\{x(t): t\in(a,b)\}\subset S$ and $f(x)=c$ for any $x\in S$, we have

$$f(x(t))=c, \quad \forall t\in(a,b)$$

$$\frac{df(x(t))}{dt}=0, \quad \forall t\in(a,b)$$

On the other hand, using the chain rule

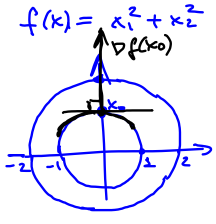
$$0=\frac{df(x(t))}{dt}=\nabla f(x(t))^T\cdot x'(t)$$

$t=t_0$: $\quad =\nabla f(x(t_0))^T\cdot x'(t_0)=\nabla f(x_0)^T x'(t_0)=0$

$\Rightarrow \nabla f(x_0)^T$ and $x'(t_0)$ are orthogonal

The gradient of $f(x)$ at x_0 is orthogonal to any curve passing throug x_0 in the level set of $f(x)$ corresponding to the level $f(x_0)$.

$f(x)=x_1^2+x_2^2$



$\nabla f(x)=2\begin{bmatrix}x_1\\x_2\end{bmatrix}$

$x_0=\begin{bmatrix}0\\1\end{bmatrix}$

$\nabla f(x_0)=\begin{bmatrix}0\\2\end{bmatrix}$

In gradient methods for minimization steps are taken in the direction opposite to the gradient.

1) x_0, the starting guess
2) $x_{k+1}=x_k+\alpha_k d_k$, $k\geq 0$
3) stopping criteria:
$\|x_{k+1}-x_k\|<\varepsilon$
or $\|\nabla f(x_k)\|<\delta$

Gradient methods: $d_k=-\nabla f(x_k)$

α_k can be chosed in several different ways

The Cauchy - Schwartz inequality: $u, v \in \mathbb{R}^n$

$$u^T v \le \|u\| \cdot \|v\| \quad \text{"="} \text{ iff } u = \lambda v.$$

$\nabla f(x_k)^T d$ – the rate of increase of f at x_k in the direction d.

$$\nabla f(x_k)^T d \le \|\nabla f(x_k)\| \cdot \|d\|, \quad \text{"="} \Leftrightarrow d = \lambda \nabla f(x_k)$$

$\Rightarrow$ the largest rate of increase is achieved when $\underline{\underline{d = \lambda \nabla f(x_k)}}$.

$-\nabla f(x_k), d$: $\quad -\nabla f(x_k)^T d \le \|\nabla f(x_k)\| \cdot \|d\|$

$$\Leftrightarrow \nabla f(x_k)^T d \ge -\|\nabla f(x_k)\| \cdot \|d\|$$

"=" $\Leftrightarrow d = -\lambda \nabla f(x_k) \Rightarrow$ the highest rate of decrease is achieved when $d = -\lambda \nabla f(x_k)$.