Lecture 3: Methods for local unconstrained optimization. Linesearch methods

Coralia Cartis, Mathematical Institute, University of Oxford

C6.2/B2: Continuous Optimization

Methods for local unconstrained optimization

minimize $f(x)$ subject to $x \in \mathbb{R}^n$ (UP) $[f \in \mathcal{C}^1(\mathbb{R}^n)$ or $f \in \mathcal{C}^2(\mathbb{R}^n)]$

A Generic Method (GM)

Choose $\epsilon > 0$ and $x^0 \in \mathbb{R}^n$.
While (TERMINATION CRITERIA not achieved), REPEAT:

- compute the change

  $$x^{k+1} - x^k = F(x^k, \texttt{problem data}), \qquad \text{[linesearch, trust-region]}$$

  to ensure $f(x^{k+1}) < f(x^k)$.

- set $x^{k+1} := x^k + F(x^k, \texttt{prob. data})$, $k := k+1$. $\square$

Methods for local unconstrained optimization

minimize $f(x)$ subject to $x \in \mathbb{R}^n$ (UP) $[f \in \mathcal{C}^1(\mathbb{R}^n)$ or $f \in \mathcal{C}^2(\mathbb{R}^n)]$

A Generic Method (GM)

Choose $\epsilon > 0$ and $x^0 \in \mathbb{R}^n$.

While (TERMINATION CRITERIA not achieved), REPEAT:

- compute the change

  $$x^{k+1} - x^k = F(x^k, \texttt{problem data}), \qquad \text{[linesearch, trust-region]}$$

  to ensure $f(x^{k+1}) < f(x^k)$.

- set $x^{k+1} := x^k + F(x^k, \texttt{prob. data})$, $k := k + 1$. $\square$

- TC: $\|\nabla f(x^k)\| \leq \epsilon$; maybe also, $\lambda_{\min}(\nabla^2 f(x^k)) \geq -\epsilon$.

Methods for local unconstrained optimization

minimize $f(x)$ subject to $x \in \mathbb{R}^n$ (UP) $[f \in \mathcal{C}^1(\mathbb{R}^n)$ or $f \in \mathcal{C}^2(\mathbb{R}^n)]$

A Generic Method (GM)

```
Choose ε > 0 and x^0 ∈ ℝ^n.
While (TERMINATION CRITERIA not achieved), REPEAT:
```

- compute the change

$$x^{k+1} - x^k = F(x^k, \texttt{problem data}),$$ [linesearch, trust-region]

to ensure $f(x^{k+1}) < f(x^k)$.

- set $x^{k+1} := x^k + F(x^k, \texttt{prob. data})$, $k := k+1$. $\square$

- TC: $\|\nabla f(x^k)\| \leq \epsilon$; maybe also, $\lambda_{\min}(\nabla^2 f(x^k)) \geq -\epsilon$.
- e.g., $x^{k+1} \equiv$ minimizer of some (simple) model of f around x^k
$\longrightarrow$ linesearch, trust-region methods.
- if $F = F(x_k, x_{k-1}, \text{problem data}) \longrightarrow$ conjugate gradients mthd.

Issues to consider about GM

Global convergence of GM:

if $\epsilon := 0$ and <u>any</u> $x^0 \in \mathbb{R}^n$: $\nabla f(x^k) \to 0$, as $k \to \infty$?

[maybe also, $\liminf_{k\to\infty} \lambda_{\min}(\nabla^2 f(x^k)) \geq 0$?]

Local convergence of GM:

if $\epsilon := 0$ and x^0 <u>sufficiently close</u> to $x^* \equiv$ stationary/local minimizer of f: $x^k \to x^*$, $k \to \infty$?

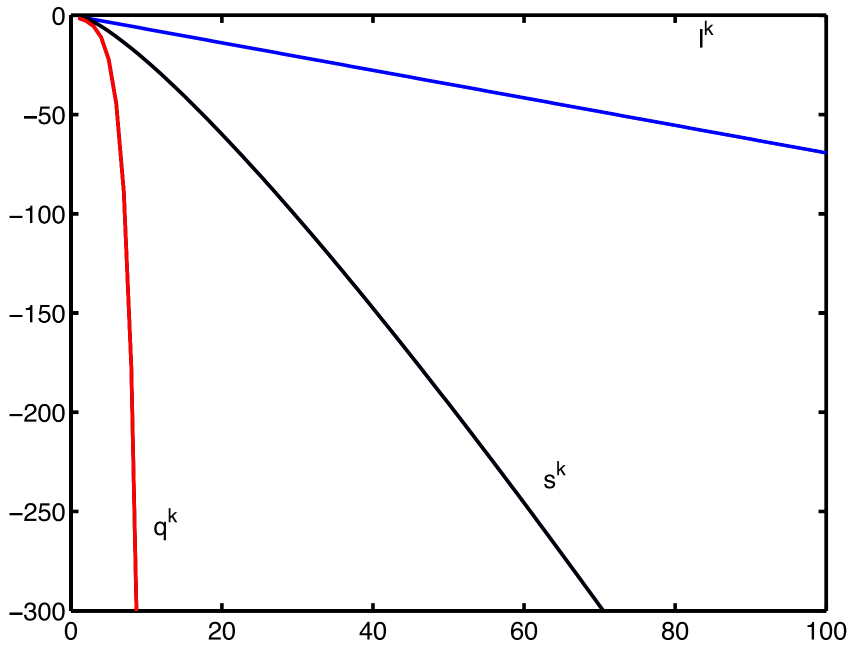
Global/local complexity of GM: count number of iterations and their cost required by GM to generate x^k within desired accuracy $\epsilon > 0$, e.g., such that $\|\nabla f(x^k)\| \leq \epsilon$.

[connection to convergence and its rate]

Rate of global/local convergence of GM.

Rates of convergence of sequences: an example

$l^k := (1/2)^k \longrightarrow 0$ linearly,
$q^k := (1/2)^{2^k} \longrightarrow 0$ quadratically,
$s^k := k^{-k} \longrightarrow 0$ superlinearly as $k \longrightarrow \infty$.



Rates of convergence on a log scale.

k	l^k	q^k
0	$\mathbf{1}$	$\mathbf{0.5}$
1	$\mathbf{0.5}$	$\mathbf{0.25}$
2	$\mathbf{0.25}$	$\mathbf{0.6 \cdot (-1)}$
3	$\mathbf{0.12}$	$\mathbf{0.4 \cdot (-2)}$
4	$\mathbf{0.6 \cdot (-2)}$	$\mathbf{0.1 \cdot (-4)}$
5	$\mathbf{0.3 \cdot (-2)}$	$\mathbf{0.2 \cdot (-9)}$
6	$\mathbf{0.2 \cdot (-2)}$	$\mathbf{0.5 \cdot (-19)}$

Notation: $(-i) := 10^{-i}$.

Rates of convergence of sequences

$\{x^k\} \subset \mathbb{R}^n, x^* \in \mathbb{R}^n; \quad x^k \to x^*$ as $k \to \infty$.
p-Rate of convergence: $x^k \to x^*$ with rate $p \geq 1$ if $\exists \rho > 0$ and $k_0 \geq 0$ such that

$$\|x^{k+1} - x^*\| \leq \rho \|x^k - x^*\|^p, \quad \forall k \geq k_0.$$

- ρ convergence factor; $e^k := x^k - x^*$ error in $x^k \approx x^*$.

Linear convergence: $p = 1 \quad \Rightarrow \rho < 1$; (asymptotically,) no of correct digits grows linearly in the number of iterations.

Quadratic convergence: $p = 2$; (asymptotically,) no of correct digits grows exponentially in the number of iterations.

Superlinear convergence: $\|x^{k+1} - x^*\| / \|x^k - x^*\| \to 0$ as $k \to \infty$. [faster than linear, slower than quadratic; practically very acceptable]

Summary: methods for local unconstrained probs.

Consider (UP), with $f \in \mathcal{C}^1$ or $\mathcal{C}^2$.

Methods:

- iterative: start from any initial 'guess' x^0, generate x^k, $k \geq 0$.
- find (approximate) local solutions, unless special structure (convexity, etc.)
- terminate when iterate within ϵ of local optimality.

Issues: global convergence, local convergence, rate of convergence, complexity.

Information employed on each iteration:

current x^k: linesearch and trust-region methods

current+previous: conjugate-gradients method etc

A generic linesearch method

(UP): minimize $f(x)$ subject to $x \in \mathbb{R}^n$, where $f \in \mathcal{C}^1$ or $\mathcal{C}^2$.

A Generic Linesearch Method (GLM)

Choose $\epsilon > 0$ and $x^0 \in \mathbb{R}^n$. For $k \geq 0$, do:
While $\|\nabla f(x^k)\| > \epsilon$, REPEAT:

- compute a descent search direction $s^k \in \mathbb{R}^n$,

$$\nabla f(x^k)^T s^k < 0;$$

- compute a stepsize $\alpha^k > 0$ along s^k such that

$$f(x^k + \alpha^k s^k) < f(x^k);$$

- set $x^{k+1} := x^k + \alpha^k s^k$ and $k := k + 1$. $\square$

Recall property of descent directions (Lemma 1, Lecture 1).

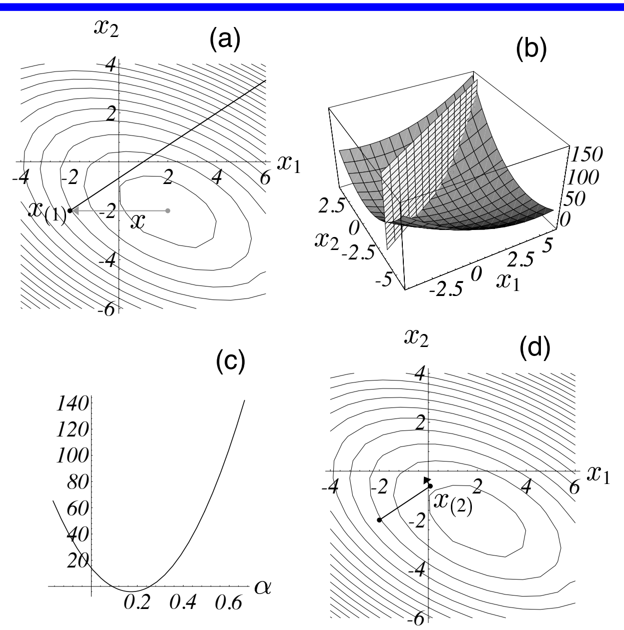
Performing a linesearch

How to compute α^k?

Exact linesearch:

$\alpha^k := \arg\min_{\alpha>0} f(x^k + \alpha s^k)$.

- computationally expensive for nonlinear objectives.



Exact linesearch for quadratic functions

Example: $q(x) = \frac{1}{2}x^T \begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix} x + (-2 \quad 8)^T x$, where $x \in \mathbb{R}^2$.

Let $x^1 := (-2 \ -2)^T$ and $s^1 := -\nabla q(x^1) = (12 \ \ 8)^T$.

Figure (a): contours of q and the line $x^1 + \alpha s^1$; (b): the plane $z(\alpha) = x^1 + \alpha s^1$ is shown cutting the q-surface; (c): plot of $\phi(\alpha)$; (d): x^2 is shown and $\phi'(\alpha^*) = 0$. $\square$

Exact linesearches for quadratic objectives

$$q(x) = g^T x + \frac{1}{2} x^T H x, \quad x \in \mathbb{R}^n,$$

and let $\phi_k(\alpha) := q(x^k + \alpha s^k)$.

Exact linesearches for quadratic objectives

$$q(x) = g^T x + \frac{1}{2} x^T H x, \quad x \in \mathbb{R}^n,$$

and let $\phi_k(\alpha) := q(x^k + \alpha s^k)$. Then

$$\phi'(\alpha) = \frac{d}{d\alpha}\phi(\alpha) = \sum_{i=1}^n \frac{dx_i}{d\alpha} \cdot \frac{\partial}{\partial x_i}\phi(\alpha)$$

$$= \sum_{i=1}^n s_i^k \frac{\partial}{\partial x_i} q(x^k + \alpha s^k) = (s^k)^T \nabla q(x^k + \alpha s^k).$$

Exact linesearches for quadratic objectives

$$q(x) = g^T x + \frac{1}{2} x^T H x, \quad x \in \mathbb{R}^n,$$

and let $\phi_k(\alpha) := q(x^k + \alpha s^k)$. Then

$$\phi'(\alpha) = \frac{d}{d\alpha}\phi(\alpha) = \sum_{i=1}^n \frac{dx_i}{d\alpha} \cdot \frac{\partial}{\partial x_i}\phi(\alpha)$$

$$= \sum_{i=1}^n s_i^k \frac{\partial}{\partial x_i} q(x^k + \alpha s^k) = (s^k)^T \nabla q(x^k + \alpha s^k).$$

- $\nabla q(x) = g + Hx$ and $\nabla q(x^k + \alpha s^k) = g + H(x^k + \alpha s^k)$.

$\Longrightarrow \quad \phi'(\alpha) = (s^k)^T \nabla q(x^k) + \alpha (s^k)^T H s^k.$

Exact linesearches for quadratic objectives

$$q(x) = g^T x + \frac{1}{2} x^T H x, \quad x \in \mathbb{R}^n,$$

and let $\phi_k(\alpha) := q(x^k + \alpha s^k)$. Then

$$\phi'(\alpha) = \frac{d}{d\alpha}\phi(\alpha) = \sum_{i=1}^n \frac{dx_i}{d\alpha} \cdot \frac{\partial}{\partial x_i}\phi(\alpha)$$

$$= \sum_{i=1}^n s_i^k \frac{\partial}{\partial x_i} q(x^k + \alpha s^k) = (s^k)^T \nabla q(x^k + \alpha s^k).$$

- $\nabla q(x) = g + Hx$ and $\nabla q(x^k + \alpha s^k) = g + H(x^k + \alpha s^k)$.

$\Longrightarrow \quad \phi'(\alpha) = (s^k)^T \nabla q(x^k) + \alpha (s^k)^T H s^k.$

Thus α^* stationary point of $\phi(\alpha)$ iff $(s^k)^T H s^k \neq 0$ and $\phi'(\alpha^*) = 0 \Longrightarrow \quad \alpha^* = -(s^k)^T \nabla q(x^k)/(s^k)^T H s^k.$

- α^* global minimizer of $\phi(\alpha)$ if $(s^k)^T H s^k > 0$.

Exact linesearches for quadratic objectives

$$q(x) = g^T x + \frac{1}{2} x^T H x, \quad x \in \mathbb{R}^n,$$

and let $\phi_k(\alpha) := q(x^k + \alpha s^k)$. Then

$$\phi'(\alpha) = \frac{d}{d\alpha}\phi(\alpha) = \sum_{i=1}^n \frac{dx_i}{d\alpha} \cdot \frac{\partial}{\partial x_i}\phi(\alpha)$$
$$= \sum_{i=1}^n s_i^k \frac{\partial}{\partial x_i} q(x^k + \alpha s^k) = (s^k)^T \nabla q(x^k + \alpha s^k).$$

- $\nabla q(x) = g + Hx$ and $\nabla q(x^k + \alpha s^k) = g + H(x^k + \alpha s^k)$.

$$\Longrightarrow \quad \phi'(\alpha) = (s^k)^T \nabla q(x^k) + \alpha (s^k)^T H s^k.$$

Thus α^* stationary point of $\phi(\alpha)$ iff $(s^k)^T H s^k \neq 0$ and $\phi'(\alpha^*) = 0 \Longrightarrow \quad \alpha^* = -(s^k)^T \nabla q(x^k)/(s^k)^T H s^k.$

- α^* global minimizer of $\phi(\alpha)$ if $(s^k)^T H s^k > 0$.
- for general f, no explicit expression of α^k; approximate minimizers of $f(x^k + \alpha s^k)$ may be used instead. [see Pb Sheet 1]