Math 408A
Line Search Methods

The Backtracking Line Search

One Dimensional Optimization and Line Search Methods

Line Search Methods

Let $f : \mathbb{R}^n \to \mathbb{R}$ be given and suppose that x_c is our current best estimate of a solution to

$$\mathcal{P} \quad \min_{x \in \mathbb{R}^n} f(x) \; .$$

Line Search Methods

Let $f : \mathbb{R}^n \to \mathbb{R}$ be given and suppose that x_c is our current best estimate of a solution to

$$\mathcal{P} \quad \min_{x \in \mathbb{R}^n} f(x) \ .$$

Given $d \in \mathbb{R}^n$, we construct the one dimensional function

$$\phi(t) := f(x_c + td) \ .$$

We can then try to approximately minimize ϕ.

Line Search Methods

Let $f : \mathbb{R}^n \to \mathbb{R}$ be given and suppose that x_c is our current best estimate of a solution to

$$\mathcal{P} \quad \min_{x \in \mathbb{R}^n} f(x) \ .$$

Given $d \in \mathbb{R}^n$, we construct the one dimensional function

$$\phi(t) := f(x_c + td) \ .$$

We can then try to approximately minimize ϕ.
We call d a **search direction** and the approximate solution $\bar{t}$ the **stepsize** or **step length.**

Line Search Methods

Let $f : \mathbb{R}^n \to \mathbb{R}$ be given and suppose that x_c is our current best estimate of a solution to

$$\mathcal{P} \quad \min_{x \in \mathbb{R}^n} f(x) \; .$$

Given $d \in \mathbb{R}^n$, we construct the one dimensional function

$$\phi(t) := f(x_c + td) \; .$$

We can then try to approximately minimize ϕ.
We call d a **search direction** and the approximate solution $\bar{t}$ the **stepsize** or **step length.**
The new estimate of a solution to $\mathcal{P}$ is

$$x_+ = x_c + \bar{t}d \; .$$

Line Search Methods

Let $f : \mathbb{R}^n \to \mathbb{R}$ be given and suppose that x_c is our current best estimate of a solution to

$$\mathcal{P} \quad \min_{x \in \mathbb{R}^n} f(x) \ .$$

Given $d \in \mathbb{R}^n$, we construct the one dimensional function

$$\phi(t) := f(x_c + td) \ .$$

We can then try to approximately minimize ϕ.
We call d a **search direction** and the approximate solution $\bar{t}$ the **stepsize** or **step length.**
The new estimate of a solution to $\mathcal{P}$ is

$$x_+ = x_c + \bar{t}d \ .$$

How should the search direction and stepsize be chosen.

The Basic Backtracking Algorithm

Assume that $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable and $d \in \mathbb{R}^n$ is a direction of strict descent at x_c, i.e., $f'(x_c; d) < 0$.

The Basic Backtracking Algorithm

Assume that $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable and $d \in \mathbb{R}^n$ is a direction of strict descent at x_c, i.e., $f'(x_c; d) < 0$.

INITIALIZATION: Choose $\gamma \in (0,1)$ and $c \in (0,1)$.

The Basic Backtracking Algorithm

Assume that $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable and $d \in \mathbb{R}^n$ is a direction of strict descent at x_c, i.e., $f'(x_c; d) < 0$.

INITIALIZATION: Choose $\gamma \in (0, 1)$ and $c \in (0, 1)$.

Having x_c obtain x_+ as follows:

The Basic Backtracking Algorithm

Assume that $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable and $d \in \mathbb{R}^n$ is a direction of strict descent at x_c, i.e., $f'(x_c; d) < 0$.

INITIALIZATION: Choose $\gamma \in (0, 1)$ and $c \in (0, 1)$.

Having x_c obtain x_+ as follows:

STEP 1: Compute the backtracking stepsize

$$\begin{aligned} t^* \quad := \quad & \max \gamma^\nu \\ & \text{s.t. } \nu \in \{0, 1, 2, \ldots\} \text{ and} \\ & f(x_c + \gamma^\nu d) \le f(x_c) + c\gamma^\nu f'(x_c; d). \end{aligned}$$

The Basic Backtracking Algorithm

Assume that $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable and $d \in \mathbb{R}^n$ is a direction of strict descent at x_c, i.e., $f'(x_c; d) < 0$.

INITIALIZATION: Choose $\gamma \in (0, 1)$ and $c \in (0, 1)$.

Having x_c obtain x_+ as follows:

STEP 1: Compute the backtracking stepsize

$$\begin{aligned} t^* \quad := \quad & \max \gamma^\nu \\ & \text{s.t. } \nu \in \{0, 1, 2, \ldots\} \text{ and} \\ & f(x_c + \gamma^\nu d) \leq f(x_c) + c\gamma^\nu f'(x_c; d). \end{aligned}$$

STEP 2: Set $x_+ = x_c + t^* d$.

The Basic Backtracking Algorithm

We need to show that the backtracking line search is well-defined and finitely terminating.

The Basic Backtracking Algorithm

We need to show that the backtracking line search is well-defined and finitely terminating.
Since $f'(x_c; d) < 0$ and $0 < c < 1$, we know

$$f'(x_c; d) < cf'(x_c; d) < 0.$$

The Basic Backtracking Algorithm

We need to show that the backtracking line search is well-defined and finitely terminating.
Since $f'(x_c; d) < 0$ and $0 < c < 1$, we know

$$f'(x_c; d) < cf'(x_c; d) < 0.$$

Hence

$$f'(x_c; d) = \lim_{t \downarrow 0} \frac{f(x_c + td) - f(x_c)}{t} < cf'(x_c; d) \ .$$

The Basic Backtracking Algorithm

We need to show that the backtracking line search is well-defined and finitely terminating.

Since $f'(x_c; d) < 0$ and $0 < c < 1$, we know

$$f'(x_c; d) < cf'(x_c; d) < 0.$$

Hence

$$f'(x_c; d) = \lim_{t \downarrow 0} \frac{f(x_c + td) - f(x_c)}{t} < cf'(x_c; d) \ .$$

Therefore, there is a $\bar{t} > 0$ such that

$$\frac{f(x_c + td) - f(x_c)}{t} < cf'(x_c; d) \quad \forall \ t \in (0, \bar{t}),$$

that is

$$f(x_c + td) < f(x_c) + ctf'(x_c; d) \quad \forall \ t \in (0, \bar{t}).$$

The Basic Backtracking Algorithm

So

$$f(x_c + td) < f(x_c) + ctf'(x_c; d) \quad \forall\ t \in (0, \bar{t}).$$

The Basic Backtracking Algorithm

So

$$f(x_c + td) < f(x_c) + ctf'(x_c; d) \quad \forall \; t \in (0, \bar{t}).$$

Since $0 < \gamma < 1$, $\gamma^\nu \downarrow 0$ as $\nu \uparrow \infty$, there is a ν_0 such that $\gamma^\nu < \bar{t}$ for all $\nu \geq \nu_0$.

The Basic Backtracking Algorithm

So

$$f(x_c + td) < f(x_c) + ctf'(x_c; d) \quad \forall\ t \in (0, \bar{t}).$$

Since $0 < \gamma < 1$, $\gamma^\nu \downarrow 0$ as $\nu \uparrow \infty$, there is a ν_0 such that $\gamma^\nu < \bar{t}$ for all $\nu \geq \nu_0$.
Consequently,

$$f(x_c + \gamma^\nu d) \leq f(x_c) + c\gamma^\nu f'(x_c; d) \quad \forall\ \nu \geq \nu_0,$$

that is, the backtracking line search is finitely terminating.

Programming the Backtracking Algorithm

Pseudo-Matlab code:

$$
\left[\begin{array}{llcl}
 & f_c & = & f(x_c) \\
 & \Delta f & = & cf'(x_c; d) \\
 & \mathrm{new}f & = & f(x_c + d) \\
 & t & = & 1 \\
\texttt{while} & \mathrm{new}f & > & f_c + t\Delta f \\
 & t & = & \gamma t \\
 & \mathrm{new}f & = & f(x_c + td) \\
\texttt{endwhile} & & &
\end{array}\right.
$$

Direction Choices

There are essentially three directions of interest:

Direction Choices

There are essentially three directions of interest:

1. Steepest Descent (or Cauchy Direction):

$$d = -\nabla f(x_c)/\|\nabla f(x_c)\| \ .$$

Direction Choices

There are essentially three directions of interest:

1. Steepest Descent (or Cauchy Direction):

$$d = -\nabla f(x_c)/\|\nabla f(x_c)\| \ .$$

2. Newton Direction:

$$d = -\nabla^2 f(x_c)^{-1}\nabla f(x_c) \ .$$

Direction Choices

There are essentially three directions of interest:

1. Steepest Descent (or Cauchy Direction):

$$d = -\nabla f(x_c)/\|\nabla f(x_c)\| \ .$$

2. Newton Direction:

$$d = -\nabla^2 f(x_c)^{-1} \nabla f(x_c) \ .$$

3. Newton-Like Direction:

$$d = -H \nabla f(x_c),$$

where $H \in \mathbb{R}^{n \times n}$ is symmetric and constructed so that

$$H \approx \nabla^2 f(x_c)^{-1} \ .$$

Descent Condition

For all of these directions we have

$$f'(x_c; -H\nabla f(x_c)) = -\nabla f(x_c)^T H\nabla f(x_c).$$

Descent Condition

For all of these directions we have

$$f'(x_c; -H\nabla f(x_c)) = -\nabla f(x_c)^T H\nabla f(x_c).$$

Thus, to obtain strict descent we need,

$$0 < \nabla f(x_c)^T H\nabla f(x_c) \ .$$

Descent Condition

For all of these directions we have

$$f'(x_c; -H\nabla f(x_c)) = -\nabla f(x_c)^T H\nabla f(x_c).$$

Thus, to obtain strict descent we need,

$$0 < \nabla f(x_c)^T H\nabla f(x_c) \; .$$

This holds, in particular, when H is positive definite.

Descent Condition

For all of these directions we have

$$f'(x_c; -H\nabla f(x_c)) = -\nabla f(x_c)^T H \nabla f(x_c).$$

Thus, to obtain strict descent we need,

$$0 < \nabla f(x_c)^T H \nabla f(x_c) \ .$$

This holds, in particular, when H is positive definite.

In the case of steepest descent $H = I$ and so

$$\nabla f(x_c)^T H \nabla f(x_c) = \|\nabla f(x_c)\|_2^2.$$

Descent Condition

For all of these directions we have

$$f'(x_c; -H\nabla f(x_c)) = -\nabla f(x_c)^T H\nabla f(x_c).$$

Thus, to obtain strict descent we need,

$$0 < \nabla f(x_c)^T H\nabla f(x_c) \ .$$

This holds, in particular, when H is positive definite.

In the case of steepest descent $H = I$ and so

$$\nabla f(x_c)^T H\nabla f(x_c) = \|\nabla f(x_c)\|_2^2.$$

In all other cases, $H \approx \nabla^2 f(x_c)^{-1}$. The condition that H be pd is related to the second-order sufficient condition for optimality, a local condition.