

Numerical Optimisation: Line search methods

Assignment Project Exam Help

<https://eduassistpro.github.io>

`f.rullan@cs.ucl.ac.uk`

Add WeChat edu_assist_pro

Department of Computer Science
Centre for Medical Image Computing
Centre for Inverse Problems
University College London

Lecture 2 & 3

Descent direction is a vector $p \in \mathbb{R}^n$ for which the function decreases.

Assignment Project Exam Help

From Taylor's theorem

$$f(x_k + tp) - f(x_k) \approx \nabla f(x_k)^T p t, \quad t \in (0, \alpha)$$

< 0

Add WeChat edu_assist_pr

Thus for $\alpha > 0$ small enough, $f(x_k + tp) < f(x_k)$

$$p^T \nabla f(x_k) = \|p\| \|\nabla f(x_k)\| \cos \theta < 0 \Leftrightarrow |\theta| > \pi/2,$$

where θ is the angle between p and $\nabla f(x_k)$.

Steepest descent direction

Steepest descent direction p

$$\min_p p^T \nabla f(x_k), \quad \text{subject to } \|p\| = 1.$$

Assignment Project Exam Help

$$\min_p p^T \nabla f(x_k) = \min_p \|p\| \|\nabla f(x_k)\| \cos(\theta) = -\|\nabla f(x_k)\|,$$

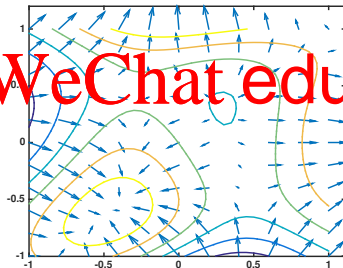
attai

re θ is

the ar

<https://eduassistpro.github.io>

Add WeChat edu_assist_pr



Newton direction

Consider the second order Taylor polynomial

$$f(x_k + p) \approx f(x_k) + p^T \nabla f(x_k) + \frac{1}{2} p^T \nabla^2 f(x_k) p =: m_2(p)$$

and assume $\nabla^2 f(x_k)$ is positive definite.

New

m_2 . Set

yields

The Newton direction is reliable when $m_2(p)$ is a close approximation to $f(x_k + p)$ i.e. $\nabla^2 f(x_k + tp)$, $t \in (0, 1)$ and $\nabla^2 f(x_k)$ are close. This is the case if $\nabla^2 f$ is sufficiently smooth and the difference is of order $\mathcal{O}(\|p\|^3)$.

$$p^T \nabla f(x_k) = -p^T \nabla^2 f(x_k) p \leq -\sigma \|p\|^2$$

for some $\sigma > 0$. Thus unless $\nabla f(x_k) = 0$ (and hence $p = 0$),

$p^T \nabla f(x_k) < 0$ and p is a descent direction.

The step length 1 is optimal for $f(x + p) = m_2(p)$, thus 1 is used unless

If ∇

be defined: if $\nabla^2 f(x_k)$ is singular, $\nabla^2 f$ 1

Otherwise, p may not be a descent direction wh remedied.

Fast local convergence (quadratic) close to the solution.

Computing the Hessian is expensive.

Quasi-Newton direction

Use symmetric positive definite (s.p.d.) approximation B_k to the Hessian $\nabla^2 f(x_k)$ in the Newton step

Assignment Project Exam Help

such that superlinear convergence is retained.

B_k is
infor

the fact that changes in gradient provide information about the second derivative of f along the search dire

Secant equation

$$\nabla f(x_k + p) = \nabla f(x_k) + B_k p$$

This equation is underdetermined, different methods quasi-Newton methods differ in the way they solve it.

Assignment Project Exam Help

Given the search direction p , the optimal reduction of the f amounts to minimising the function of one variable

This is
inex

<https://eduassistpro.github.io>

Choice of step size is important. Too small steps mean convergence, too large steps may not lead to reduction of objective function f .

Add WeChat: edu_assist_pro

Conditions for decrease

Simple condition: require $f(x_k + \alpha p) < f(x_k)$.

Consider a sequence $f(x_k) = 5/k, k = 1, 2, \dots$. This sequence is decreasing but its limiting value is 0, while the minimum of a convex function can be smaller than 0.

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu_assist_pr

Figure: Nocedal Wright Fig 3.2

The decrease is insufficient to converge to the minimum of a convex function. Hence we need conditions for sufficient decrease.

Sufficient decrease condition

Armijo condition

$$f(x_k + \alpha p) \leq f(x_k) + c_1 \alpha p^T \nabla f(x_k) =: \ell(\alpha),$$

for some $c_1 \in (0, 1)$ [Typically small, $c_1 = 10^{-4}$]

$\ell(\alpha)$ is a linear function with negative slope $c_1 p^T \nabla f(x_k) < 0$,

$\ell(\alpha) = f(x_k) + \alpha \phi'(0)$.

From the Armijo condition, $\ell(\alpha) \leq f(x_k + \alpha p)$ for all $\alpha \in (0, \alpha_k)$,
thus f decreases along the search direction p .

Figure: Nocedal Wright Fig 3.3

Curvature condition

Armijo condition is satisfied for all sufficiently small α , so we need another condition to avoid very small steps.

Curvature condition

$$\underline{p^T} \quad \underline{f(x_k + \alpha p)} \quad \underline{c_2} \quad \underline{p^T} \quad \underline{f(x_k)}, \quad c_2 \in (c_1, 1).$$

<https://eduassistpro.github.io>

- direction p .
- If $\phi'(\alpha)$ is strongly negative, there is a good p significant decrease along p .
- If $\phi'(\alpha)$ is slightly negative (or even positive) prospect of little decrease and hence we terminate the line search.
- Typically $c_2 = 0.9$ for a Newton or quasi Newton direction, $c_2 = 0.1$ for nonlinear conjugate gradient.

Curvature condition

$$\underbrace{p^T \nabla f(x_k + \alpha p)}_{d(x_k)} \geq c_2 \underbrace{p^T \nabla f(x_k)}_{d(x_k)}, \quad c_2 \in (c_1, 1).$$

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu_assist_pr

Figure: Nocedal Wright Fig 3.4

The sufficient decrease (Armijo rule) and curvature conditions together are called **Wolfe conditions**

Assignment Project Exam Help

$$\begin{aligned} f(x_k + \alpha p) &\leq f(x_k) + c_1 \alpha p^T \nabla f(x_k), \\ p^T \nabla f(x_k + \alpha p) &\geq c_2 p^T \nabla f(x_k), \end{aligned}$$

for 0

<https://eduassistpro.github.io>

Poss

strong Wolfe conditions to disallow “to

$\phi'(\alpha)$

Add WeChat edu_assist_pr

$$\begin{aligned} f(x_k + \alpha p) &\leq f(x_k) + c_1 \alpha p^T \nabla f(x_k), \\ |p^T \nabla f(x_k + \alpha p)| &\leq c_2 |p^T \nabla f(x_k)|, \end{aligned}$$

for $0 < c_1 < c_2 < 1$.

Wolfe conditions: existence

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ continuously differentiable and p be a descent direction at x_k . If f is bounded below along the ray $\{x_k + \alpha p \mid \alpha \geq 0\}$, then there exists an interval of step lengths satisfying both the Wolfe conditions and strong Wolfe conditions.

Pro

For

$\ell(\alpha)$

$c_1 p^T \nabla f(x_k) < 0$ and for small α , $\ell(\alpha)$

$\ell(\alpha)$ has to intersect $\phi(\alpha)$ at least once. Let

value for which

$$\phi(\alpha') = f(x_k + \alpha' p) = f(x_k) + \alpha' c_1 p^T \nabla f(x_k) = \ell(\alpha').$$

Then the sufficient decrease condition holds for all $\alpha \leq \alpha'$.

Furthermore, by the mean value theorem

$$\exists \alpha'' \in (0, \alpha') : f(x_k + \alpha' p) - f(x_k) = \alpha' p^T \nabla f(x_k + \alpha'' p)$$

and we obtain

$$p^T \nabla f(x_k + \alpha'' p) = c_1 p^T \nabla f(x_k) > c_2 p^T \nabla f(x_k)$$

since

cond
stric

s Wolfe

holds

conditions) also holds in an interval containing

as all terms in the last equation are negative strong W
conditions hold for the same interval

Wolfe conditions are scale-invariant in the sense that are unaffected by scaling the function or affine change of variables. They can be used in most line search methods and are particularly important for quasi-Newton methods.

$$f(x_k) + (1 - c)\alpha p^T \nabla f(x_k) \leq f(x_k + \alpha p) \leq f(x_k) + c\alpha p^T \nabla f(x_k)$$

with $0 \leq c \leq 1/2$

The second inequality is the sufficient decrease condition. The first inequality controls the step length from below. Disadvantage w.r.t ϕ .

<https://eduassistpro.github.io>

Add WeChat edu_assist_pr

Figure: Nocedal Wright Fig 3.6

Backtracking line search

1: Choose $\bar{\alpha} > 0, \rho \in (0, 1), c \in (0, 1)$

2: Set $\alpha = \bar{\alpha}$

3: repeat

4: $\alpha = \rho\alpha$

5:

T

- <https://eduassistpro.github.io>
- Prevents too short step lengths: the accept factor ρ of the previous value, α , violating the sufficient decrease condition
- ρ can vary in $[\rho_{\min}, \rho_{\max}] \subset (0, 1)$
- In Newton and quasi-Newton methods $\bar{\alpha} = 1$, but different values can be appropriate for other algorithms.
- Well suited for Newton methods, less appropriate for quasi-Newton and conjugate gradient methods.

Consider an iteration

$$x_{k+1} = x_k + \alpha_k p_k, \quad k = 0, 1, \dots,$$

where p_k is a descent direction and α_k satisfied the Wolfe cond

Let f

opens

If ∇f

<https://eduassistpro.github.io>

Add WeChat edu_assist_pro

then

$$\sum_{k \geq 0} \cos^2 \theta_k \|\nabla f(x_k)\|^2 < \infty,$$

where $\theta_k = \angle(p_k, -\nabla f(x_k))$.

Subtracting $p_k^T \nabla f(x_k)$ from both sides of curvature condition

$$p_k^T \nabla f(\underbrace{x_k + \alpha p_k}_{=x_{k+1}}) \geq c_2 p_k^T \nabla f(x_k)$$

Assignment Project Exam Help

we obtain

<https://eduassistpro.github.io>

On the other hand the Lipschitz condition implies

$$p_k^T (\nabla f(x_{k+1}) - \nabla f(x_k)) \leq \|\nabla f(x_{k+1}) - \nabla f(x_k)\| \|p_k\| \leq L \alpha_k^2.$$

Combining the two inequalities we obtain a lower bound on the step size

$$\alpha_k \geq \frac{c_2 - 1}{L} \frac{p_k^T \nabla f(x_k)}{\|p_k\|^2}.$$

Substituting this inequality into the sufficient decrease condition

$$f(x_{k+1}) \leq f(x_k) + c_1 \frac{c_2 - 1}{L} \frac{(p_k^T \nabla f(x_k))^2}{\|p_k\|^2}$$

with $\cos \theta_k = -\frac{p_k^T \nabla f(x_k)}{\|\nabla f(x_k)\| \|p_k\|}$ yields

$$f(x_{k+1}) \leq f(x_k) - c \cos^2 \theta_k \|\nabla f(x_k)\|^2,$$

where

Sum

$$f(x_{k+1}) \leq f(x_0) - c \sum_{j=0}^k \cos^2 \theta_j \|\nabla f(x_j)\|^2$$

and since f is bounded from below

$$\sum_{j=0}^k \cos^2 \theta_j \|\nabla f(x_j)\|^2 \leq (f(x_0) - f(x_{k+1}))/c < C$$

where $C > 0$ is some positive constant. Taking limits

$$\sum_{k=0}^{\infty} \cos^2 \theta_k \|\nabla f(x_k)\|^2 < \infty.$$

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu_assist_pro

Global convergence

Goldstein or strong Wolfe conditions also imply the Zoutendijk condition

$$\sum_{k=0}^{\infty} \cos^2 \theta_k \|\nabla f(x_k)\|^2 < \infty.$$

Assignment Project Exam Help

The Zoutendijk condition implies

which <https://eduassistpro.github.io>
search algorithms.

If the method ensures that $\cos \theta_k \geq \delta >$
away from $\pm \frac{\pi}{2}$, it follows that

$$\lim_{k \rightarrow \infty} \|\nabla f(x_k)\| = 0.$$

This is the strongest global convergence result that can be obtained for such iteration (convergence to a stationary point) without additional assumptions.

In particular, the steepest descent ($p_k = -\nabla f(x_k)$) produces a gradient sequence which converges to 0 if it uses a line search satisfying Wolfe or Goldstein conditions.

For so
only a

<https://eduassistpro.github.io>

$$k \rightarrow \infty \quad k$$

i.e. only subsequence of gradient norms
rather than the whole sequence.

Add WeChat edu_assist_pro

Those limits can be proved by contradiction:

Suppose that $\|\nabla f(x_k)\| \geq \gamma$ for some $\gamma > 0$ for all k sufficiently large. Then from $\cos^2 \theta_k \|\nabla f(x_k)\|^2 \rightarrow 0$ we conclude that $\cos \theta_k \rightarrow 0$ i.e. the entire sequence $\{\cos \theta_k\}$ converges to 0.

Thus to show the weak convergence result it is enough to show that a subsequence $\cos \theta_{k_j}$ is bounded away from 0.

Cons

- (i) dec
- (ii) ev

satisfying the Wolfe or Goldstein conditions.

Since $\cos \theta_k \neq 1$ for steepest descent steps, this p subsequence bounded away from 0. The algorithm something better in remaining $m - 1$ iterates, while the occasional steepest descent step will guarantee the overall (weak) global convergence.

Unfortunately, rapid convergence sometimes conflicts with global convergence.

Exa

step i

while

solution, the Newton step may not even be a descent direction far away from the solution.

The challenge: design algorithms with good global properties and rapid convergence rate.

Steepest descent with exact line search for strictly convex quadratic function

Assignment Project Exam Help

$$f(x) = \frac{1}{2}x^T Q x - l^T x$$

where

<https://eduassistpro.github.io>

Add WeChat edu_assist_pr

Figure: Nocedal Wright Fig 3.7

Steepest descent

Characteristic zig-zag due to elongated shape of the ellipse. If the level sets were circles instead, the steepest descent would need one step only.

Assignment Project Exam Help

Convergence rate of steepest descent with exact line search:

<https://eduassistpro.github.io>

where $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ are the ei

$\|x\|_Q^2 = x^T Q x$

Note: for quadratic strictly con
obtain objective function and (for free) iterate co

The objective function convergence rate is essentially the same for steepest descent with exact line search when applied to a twice continuously differentiable nonlinear function satisfying sufficient conditions at x^* .

Local convergence rate: Newton methods

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be twice continuously differentiable with Lipschitz continuous Hessian in a neighbourhood of the solution x^* satisfying the sufficient conditions. Note that the Hessian $\nabla^2 f$ is positive definite also in the vicinity of the solution x^* .

The iterates x computed by the Newton method (note step length 1)

<https://eduassistpro.github.io>

converge locally quadratically i.e. for starting point x_0 sufficiently close to x^* .

The sequence of gradient norms $\|\nabla f_k\|$ converges quadratically to 0.

Local convergence: note that away from the solution $\nabla^2 f_k$ may not be positive definite and hence p_k may not be a descent direction. Global convergence with Hessian modification is discussed later.

Convergence rates: Newton-type methods

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be twice continuously differentiable.

Let $\{x_k\}$ be a sequence generated by a descent method

Assignment Project Exam Help

for step sizes satisfying Wolfe conditions with $c_1 \leq 1/2$.

If the se

suffi

<https://eduassistpro.github.io>

$$\lim_{k \rightarrow \infty} \frac{\|p_k\|}{\|x_k - x^*\|} = 0$$

then for all $k > k_0$, the step length α_k

that choice of $\alpha_k = 1$, $k > k_0$, the sequen

superlinearly.

Note: once close enough to the solution so that $\nabla^2 f(x_k)$ became s.p.d., the limit is trivially satisfied and for $\alpha_k = 1$ we recover local quadratic convergence.

Convergence rates: quasi-Newton methods

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be twice continuously differentiable.

Let $\{x_k\}$ be a sequence generated by a quasi-Newton method
(note step length 1, B_k s.p.d.)

Assignment Project Exam Help

$$x_{k+1} = x_k - \underline{B_k^{-1}} \underline{\nabla f(x_k)}.$$

Assu
suffi
if

<https://eduassistpro.github.io>

Add WeChat edu_assist_pr

$$\lim_{k \rightarrow \infty} \frac{\|(B_k - \nabla^2 f(x^*))\|}{\|p_k\|}$$

Note: the superlinear convergence rate can be attained even if the sequence $\{B_k\}$ does not converge to $\nabla^2 f(x^*)$. It suffices that B_k becomes increasingly accurate approximation to $\nabla^2 f(x^*)$ along the search direction p_k . Quasi-Newton methods use it to construct B_k .

Away from the solution, the Hessian may not be positive definite, and the Newton direction may not be a descent direction. The general solution is to consider positive definite approximations.

B_k is
sufficiently

Global convergence results can be established for Newton method with Hessian modification and step satisfying W
Armijo backtracking conditions provided that

$\kappa(B_k) = \|B_k\| \|B_k^{-1}\| \leq C$ for some $C > 0$ and all k whenever the sequence of the Hessians $\{\nabla^2 f(x_k)\}$ is bounded.

Eigenvalue decomposition

$$\nabla^2 f(x_k) = Q \Lambda Q^T = \sum_{i=1}^n \lambda_i q_i q_i^T.$$

Exa

<https://eduassistpro.github.io>

Add WeChat [edu_assist_pro](#)

The Newton step: $p_k = (-0.1, 1, 2)^T$

As $p_k^T \nabla f(x_k) > 0$, it is not a descent direction.

Eigenvalue modifications (not practical):

Replace all negative eigenvalues with $\delta = \sqrt{\mathbf{u}} = 10^{-8}$, where $\mathbf{u} = 10^{-16}$ is the machine precision.

$$B_k = \sum_{i=1}^2 \lambda_i q_i q_i^T + \delta q_3 q_3^T = \text{diag}(10, 3, 10^{-8})$$

B_k is
direc

e

<https://eduassistpro.github.io>

$$p_k = -B_k^{-1} \nabla f_k = - \sum_{i=1}^2 \frac{1}{\lambda_i} q_i q_i^T \nabla f_k - -) q_3.$$

p_k is a descent direction but the length is very large, not in line with local validity of the Newton approximation. Thus p_k may be ineffective.

Adapt choice of δ to avoid excessive lengths. Even $\delta = 0$ which eliminates direction q_3 .

Let A is symmetric $A = Q\Lambda Q^T$.

The correction matrix ΔA of minimum Frobenius norm that ensures $\lambda_{\min}(A + \Delta A) \geq \delta$ is given by

$$\Delta A = Q \operatorname{diag}(\tau); Q^T, \text{ with } \tau_i = \begin{cases} 0, & \lambda_i \geq \delta \\ \delta - \lambda_i, & \lambda_i < \delta. \end{cases}$$

and th

<https://eduassistpro.github.io>

Frobenius norm is defined $\|A\|_F^2 = \sum_{i=1}^n$

The correction matrix ΔA of minimum Frobenius norm that satisfies $\lambda_{\min}(A + \Delta A) \geq \delta$ is given by

$$\Delta A = \tau I, \quad \text{with } \tau = \max(0, \delta - \lambda_{\min}(A)).$$

and the modified matrix has the form $A + \tau I$.

Assignment Project Exam Help

Simple idea

If min

(e.g. 1

Atte

<https://eduassistpro.github.io>

If not successful, increase $\tau_{k+1} = \max(\tau_k, \lambda_{\min}(A) + \epsilon)$

Add WeChat edu_assist_pro

Drawback: possibly multiple failed attempts to find $\lambda_{\min}(A)$

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu_assist_pr

Figure: Nocedal Wright Ex. 3.1

For A indefinite:

- The factorisation $A = LDL^T$ may not exist.
- Even if it does exist, the algorithm can be unstable

- <https://eduassistpro.github.io>

A .

- Instead, modify A during the factorisation
elements of D are sufficiently positive
and D are not too large.

Modified Cholesky decomposition

Choose $\delta, \beta > 0$. While computing j th column of L, D ensure

$$d_j \geq \delta, \quad |m_{ij}| \leq \beta, \quad i = j+1, j+2, \dots, n,$$

where $m_{ij} = l_{ij} \sqrt{d_j}$.

Assignment Project Exam Help

To satisfy these bounds we only need to change how d_j is computed

<https://eduassistpro.github.io>

where $c_{ij} = l_{ij} d_j$. Note: θ_j can be computed by computing c_{ij} , $i < j$ only needs previous

Add WeChat edu_assist_pro

Verification:

$d_j \geq \delta$ due to taking maximum

$$|m_{ij}| = |l_{ij} \sqrt{d_j}| = \frac{|c_{ij}|}{\sqrt{d_j}} \leq \frac{|c_{ij}| \beta}{\theta_j} \leq \beta, \quad \forall i > j.$$

Modified Cholesky decomposition

Properties:

- Modifies the Hessian during factorization where necessary.
- The modified Cholesky factors exist and are bounded relative to the norm of the actual Hessian.
- It does not modify Hessian if it is sufficiently positive definite.

This is

intro

size of t

$$PAP^T + E = LDL^T$$

where E is a nonnegative diagonal matrix that I sufficiently positive definite.

It has been shown, that the matrices obtained by this modified Cholesky algorithm to the exact Hessian $\nabla^2 f(x_k)$ have bounded condition numbers, hence some global convergence results can be obtained.

Step length selection

How to find a step length satisfying one of the termination conditions e.g. Wolfe etc. for

Assignment Project Exam Help

$$\phi(\alpha) = f(x_k + \alpha p_k),$$

where

If f is

global minimiser along the ray $x_k + \alpha p$ which can be calculated analytically

Add WeChat

$$\alpha_k = -\frac{p_k^T \nabla f}{p_k^T Q}$$

For general nonlinear functions iterative approach is necessary.

Line search algorithms can be classified according to the information they use:

Assignment Project Exam Help

Methods using only function evaluations can be very inefficient as they need to continue iterating until a very small interval has been found

Methods
which require gradients to evaluate.

<https://eduassistpro.github.io>

Typically, they consist of two phases:
an interval containing acceptable step lengths a
phase which locates the final step in the interval.

Add WeChat edu_assist_pro

Aim: find a step length α that satisfies sufficient decrease condition without being too small. [similarity to backtracking]

Assignment Project Exam Help

Sufficient decrease condition

<https://eduassistpro.github.io>

We want

Initial guess α_0 : check sufficient decrease condition

Add WeChat edu_assist_pro

$$\phi(\alpha_0) \leq \phi(0) + c_1 \alpha_0$$

If satisfied terminate the search. Otherwise, $[0, \alpha_0]$ contains acceptable step lengths.

Quadratic approximation $\phi_q(\alpha)$ to ϕ by interpolating the available information: $\phi_q(0) = \phi(0)$, $\phi'_q(0) = \phi'(0)$ and $\phi_q(\alpha_0) = \phi(\alpha_0)$ yields

$$\phi_q(\alpha) = \frac{\phi(\alpha_0) - \phi(0) - \alpha_0 \phi'(0)}{\alpha_0^2} \alpha^2 + \phi'(0) \alpha + \phi(0).$$

Assignment Project Exam Help

The new trial value α_1 is defined as the minimiser of ϕ_q i.e.

<https://eduassistpro.github.io>

If sufficient decrease condition is satisfied, termi

Otherwise, construct cubic interpolating the fo

$\phi_c(0) = \phi(0)$, $\phi'_c(0) = \phi'(0)$, $\phi_c(\alpha_0) = \phi(\alpha_0)$, $\phi_c(\alpha_1) = \phi(\alpha_1)$ yields

$$\phi_c(\alpha) = a\alpha^3 + b\alpha^2 + \alpha\phi'(0) + \phi(0),$$

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{\alpha_0^2 \alpha_1^2 (\alpha_1 - \alpha_0)} \begin{bmatrix} \alpha_0^2 & -\alpha_1^2 \\ -\alpha_0^3 & \alpha_1^3 \end{bmatrix} \begin{bmatrix} \phi(\alpha_1) - \phi(0) - \phi'(0)\alpha_1 \\ \phi(\alpha_0) - \phi(0) - \phi'(0)\alpha_0 \end{bmatrix}$$

By differentiating ϕ_c we find the minimiser $\alpha_2 \in [0, \alpha_1]$

$$\alpha_2 = \frac{-b + \sqrt{b^2 - 3a\phi'(0)}}{3a}$$

Assignment Project Exam Help

If necessary repeat the cubic interpolation with $\phi_c(0) = \phi(0)$,

$\phi'_c(0)$

$\phi_c(\alpha_1)$

suffi

fies the

<https://eduassistpro.github.io>

Safeguard: If any α_i is either too close to

than α_{i-1} we reset $\alpha_i = \alpha_{i-1}/2$.

Add WeChat edu_assist_pro

If derivatives can be computed along the function v

additional cost, we can also devise variant interpolating ϕ, ϕ' at two most recent values.

Assignment Project Exam Help

For Newton and quasi Newton $\alpha_0 = 1$. This ensures that unit step
length
cond

For m

steepest descent or conjugate gradient it is impor
available information to make the initial guess e.

<https://eduassistpro.github.io>

Add WeChat edu_assist_pr

- First order change in function at iterate x_k will be the same as that obtained at previous step

i.e. $\alpha_0 p_k^T \nabla f(x_k) = \alpha_{k-1} p_{k-1}^T \nabla f(x_{k-1})$

Assignment Project Exam Help

- <https://eduassistpro.github.io>

$$\alpha_0 = \frac{2(f(x_k) - \phi')}{p_k^T \nabla f(x_k)}$$

Add WeChat edu_assist_pro

It can be shown that if $x_k \rightarrow x^*$ s

converges to 1. If we adjust by setting $\alpha_0 = \min(1, 1.01\alpha_0)$ we find that the unit step length will eventually always be tried and accepted and the superlinear convergence will be observed.