Paper Number(s): **E4.29**
C1.1

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE
UNIVERSITY OF LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2001

MSc and EEE PART IV: M.Eng. and ACGI

OPTIMIZATION

Tuesday, 8 May 10:00 am

There are SIX questions on this paper.

Answer FOUR questions.

Time allowed: 3:00 hours

Examiners: Clark,J.M.C. and Allwright,J.C.

Corrected Copy

None

Special instructions for invigilators: **None**

Information for candidates:

$\|x\|$ denotes the Euclidean norm, $\sqrt{(x^T x)}$, of the vector x.

∇v denotes the gradient of v; that is the (column) vector of first-order partial derivatives of a function v on $\Re^n$;

$\nabla^2 v$ denotes the Hessian matrix of second-order partial derivatives of v .

$O(t)$ and $o(t)$ denote the Landau order symbols:

$f(t) = O(t)$ if $|f(t)|/t$ is bounded for all t sufficiently small ;

$f(t) = o(t)$ if $\lim_{t\to 0} |f(t)|/t = 0$.

(A vector function is also denoted by $O(t)$ or $o(t)$ if its components have the corresponding property).

A corollary of Taylor's theorem is that twice continuously differentiable functions v on $\Re^n$ can be expanded as follows: for $x, z \in R^n$

$$v(x+z) = v(x) + \nabla v(x)^T z + \tfrac{1}{2} z^T \nabla^2 v(x) z + o(\|x\|^2).$$

A "smooth function" is to be taken to mean a "function that possesses continuous derivatives of all relevant orders".

OptionButton1

1. (a) Suppose $v(x)$ is a smooth convex function on the plane. Prove that the point $\hat{x} = (1, \hat{x}_2)$, where $|\hat{x}_2| < 1$, is a minimizer of v, restricted to the square domain $F = \{(x_1, x_2): |x_1| \leq 1, |x_2| \leq 1\}$, if

$$\frac{\partial v}{\partial x_1}(1, \hat{x}_2) \leq 0,$$

$$\frac{\partial v}{\partial x_2}(1, \hat{x}_2) = 0 .$$

(Hint: use the convexity of v and the fact that the directional derivative $\nabla v(\hat{x})^T (y - \hat{x})$ is the limit, for ε decreasing to zero, of $(v(\hat{x}) + \varepsilon(v(y) - v(\hat{x})))/\varepsilon$ to prove that, for any $y \in F$, $v(y) - v(\hat{x}) \geq \nabla v(\hat{x})^T (y - \hat{x}))$.

(b) Consider an application of two-stage receding-horizon control design to the problem of regulating the constrained first-order system

$$y_{k+1} = y_k - u_k, \qquad |u_k| \leq 1.$$

The function $f(y)$ in the feedback law $u_k = f(y_k)$ is taken to be the first component $\hat{u}_0(y)$ of the pair $(\hat{u}_0(y), (\hat{u}_1(y))$ that minimizes, for each y, the cost function

$$v(y;\ u_0, u_1) = \tfrac{1}{2}(y - u_0)^2 + \tfrac{1}{2}(y - u_0 - u_1)^2$$

over the square $\{(u_0, u_1): |u_0| \leq 1, |u_1| \leq 1\}$.

It turns out that the saturated "dead-beat" law

$$\begin{aligned} f(y) &= 1 \text{ if } y \geq 1, \\ &= y \text{ if } |y| \leq 1, \\ &= -1 \text{ if } y \leq -1 \end{aligned}$$

fulfils the design requirements. Establish that it does so for the range of values of $y: 0 \leq y \leq 2$, using where necessary the assertion in (a). Illustrate your answer with a sketch of the range of minimizing points $(\hat{u}_0(y), \hat{u}_1(y))$ in the (u_0, u_1) plane.

2. (a) Give necessary and sufficient "second-order" conditions for a point $\hat{x} \in \Re^n$ to be an isolated local minimizer of a smooth function v on $\Re^n$. (Here, an *isolated local minimizer* refers to a point $\hat{x}$ that is a unique minimizer of v over a sufficiently small neighbourhood of itself; that is, there is a positive distance between $\hat{x}$ and any other local minimizer).

Determine the isolated local minimizers, if any, of the following functions

(*i*) $$v_1(x_1, x_2) = x_1 x_2 + \tfrac{1}{4}(x_1^4 + x_2^4),$$

(*ii*) $$v_2(x_1, x_2) = 1 + x_1^2 + 2x_1 x_2 + x_2^2$$

and justify your choices.

(b) Suppose $v : \Re^n \to \Re$ is a non-negative smooth objective function with only isolated stationary points and with bounded level sets $\{x : v(x) \leq c\}$. A steepest-descent method with Armijo line search is used to approximate a local minimizer. Describe this algorithm. Would you expect the algorithm always to converge to a local minimizer? If so, at what rate?

3. The Newton algorithm for minimizing a smooth function $v : \Re^n \to \Re$ generates approximations x_n to a local minimizer $\hat{x}$ according to the recursion
$x_{n+1} = x_n - (\nabla^2 v(x_n))^{-1} \nabla v(x_n)$.

(a) Let

$$\tilde{v}(\bar{x}; x) = v(\bar{x}) + \nabla v(\bar{x})^T (x - \bar{x}) + \tfrac{1}{2}(x - \bar{x})^T \nabla^2 v(\bar{x})(x - \bar{x})$$

be the second-order expansion of $v(x)$ about $\bar{x}$. Show that, as long as the Hessian matrix $\nabla^2 v(x_n)$ is positive definite and x_n is *not* a stationary point of v,

$$\tilde{v}(x_n; x_{n+1}) < v(x_n).$$

(b) Establish that the stationary point of

$$v(x) = x^3 - 3x + 1 \qquad x \in \Re$$

at $x = 1$ is a local minimizer. Determine the Newton algorithm in this case and calculate the first two approximations to the minimizer with $x_0 = 1.1$ taken as the initial approximation.

(c) For smooth functions v of a real variable, the sequence of steps
$s_n = x_{n+1} - x_n$
generated by the Newton algorithm possesses the property (if the s_n converge to zero) that for increasing n the ratio $\frac{s_{n+1}}{s_n^2}$ converges to a constant. Use this property and your previous calculations to obtain an estimate of the third approximation x_3 given by the algorithm described in (b).
(Hint: $\frac{s_{n+1}}{s_n^2}$ and $\frac{s_n}{s_{n-1}^2}$ converge to the same constant).
How would you describe the rate of convergence of the algorithm in this case?

4. Let L be a vector subspace of $\Re^n$ that is spanned by a collection of vectors $\{z_1, \cdots, z_m\}$; that is,

$$L = L[z_1, \cdots, z_m] = \{x \in \Re^n : \ x = a_1 z_1 + \cdots + a_m z_m \quad \text{for some } a_1, a_2, \cdots, a_m \in \Re\}.$$

For any $x_0 \in \Re^n$, the linear variety $x_0 + L$ is defined to be

$$x_0 + L = \left\{x \in \Re^n : \ x = x_0 + z \qquad \text{for some } z \in L\right\}.$$

Note that $x_1 + L$ coincides with $x_0 + L$ if $x_1 - x_0 \in L$.

(a) Prove that $\hat{x} \in x_0 + L$ is the global minimizer of a smooth convex function v constrained to the variety $x_0 + L$ if

$$\nabla v(\hat{x})^T z_i = 0 \qquad \text{for } i = 1, \cdots, m \ .$$

You may use the characterization of a minimizer of v on $x_0 + L$ as a point $\hat{x} \in x_0 + L$ for which

$$\nabla v(\hat{x})^T (x - \hat{x}) \geq 0 \qquad \text{for all } \ x \in x_0 + L.$$

(b) The response of a linear system is expressed in terms of the inputs by

$$y_{k+1} = 3u_k + 2u_{k-1} + u_{k-2}.$$

(i) Determine (column) vectors $z_1, z_2 \in \Re^3$ that form a basis for the two-dimensional vector subspace L of control triples $(u_1, u_2, u_3)^T$ that force y_4 to take the value zero.

(ii) Consider the problem of minimizing a cost

$$v(u_1, u_2, u_3) = u_1^2 + u_2^2 + u_3^2$$

subject to the terminal constraint that $y_4 = 1$. Show that the constrained set of control triples $(u_1, u_2, u_3)^T$ takes the form of a linear variety $(1,0,0)^T + L$, where L is the subspace described in (i). Determine the optimal control triple $(\hat{u}_1, \hat{u}_2, \hat{u}_3)^T$.

5. Consider a non-linear least squares problem, in which the objective function is a sum of squares of "residuals" $r_k(x)$:

$$v(x) = \tfrac{1}{2}\sum_{k=1}^{m} r_k(x)^2, \qquad\qquad x \in \Re^n .$$

(a) In the Gauss-Newton method the basic iteration that is used to generate approximations to the minimizer $\hat{x}$ of $v(x)$ depends only on evaluation of the residuals and their gradients:

$$x^+ = x - H(x)^{-1} \sum_{k=1}^{m} r_k(x)\, \nabla r_k(x), \qquad x \in \Re^d ,$$

where, here, x is the current iterate, x^+ the next iterate and

$$H(x) = \sum_{k=1}^{m} \nabla r_k(x)\, \nabla r_k(x)^T .$$

Show that x^+ coincides with the minimizer with respect to z of the squared norm of the vector of residuals linearized about x:

$$\bar{r}\,(x;\, z) = (\,\bar{r}_1(x;\, z), \ldots, \bar{r}_m(x;\, z)\,)^T ,$$

where

$$\bar{r}_k(x;\, z) = r_k(x) + \nabla r_k(x)^T (z - x) \;\text{ for }\; k = 1, 2, \cdots, m .$$

(b) The output of a discrete-time linear system is modelled by the equation

$$y_k = a + b\, p^k + c\, q^k + d_k \qquad k = 1, 2, \cdots, \; .$$

The unknown parameters a, b, c, p and q are to be estimated; the d_k are unknown disturbances that are believed to be very small or zero. A sequence $\bar{y}_k$ of outputs is measured for $k = 1, \cdots, 100$. Formulate a non-linear least squares problem the solution of which provides estimates for the unknown parameters, and obtain an expression for the gradient $\nabla v(x)$ of the objective function.

(c) Why is it appropriate to use the Gauss-Newton method rather than the full Newton method for the problem in part (b)? Comment on the likely rate of convergence of the Gauss-Newton method.

6. In a particular restricted step method that is used for the minimization over the plane of smooth functions v with indefinite Hessian, the iterates approximating the minimizer are generated as follows.

If x^c is the current iterate, the next iterate x^+ is taken to be $x^c + s^+$, where s^+ minimises the second-order approximation to $v(x)$ - $v(x^c)$:

$$\bar{v}(x^c;\ s) \;=\; \tfrac{1}{2}s^T Cs + b^T s$$

over the disc $\{s : s_1^2 + s_2^2 \le h^2\}$ of radius h. Here C is the Hessian matrix $\nabla^2 v(x^c)$ and b the gradient $\nabla v(x^c)$.

(a) Suppose that C has a negative eigenvalue. Then it can be shown that s^+ lies on the edge of the disc; that is, $s^{+T}s^+ = h^2$.

Let $$C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad s^+ = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}.$$

Using the method of Lagrange multipliers show that s^+ satisfies the equations

$$s_1^2 + s_2^2 = h^2\ ,$$

$$s_2(c_{11}s_1 + c_{12}s_2 + b_1) = s_1(c_{21}s_1 + c_{22}s_2 + b_2).$$

(Hint: eliminate the multiplier λ from the necessary conditions associated with the Lagrangian.)

(b) The quadratic equations in (a) in general have four possible solutions. Determine these solutions in an application of the restricted step method to the function

$$v(x_1, x_2) = x_1x_2 + \tfrac{1}{3}(x_2^3 - x_1^3) + \tfrac{1}{4}(x_1^4 + x_2^4)$$

where the current iterate x^c is taken to be $(0,\ 0)^T$.

(c) Devise a sensible strategy for selecting from these solutions a suitable next step $x^+ - x^c$ and so determine x^+.

Optimization 2001

1 (a) solution Take any $y \in F$. $\frac{1}{12}$

By convexity of v, for any $0 \le \varepsilon \le 1$

$$v((1-\varepsilon)\hat{x} + \varepsilon y) \leq (1-\varepsilon)v(\hat{x}) + \varepsilon v(y).$$

5 So $$v(y) - v(\hat{x}) \geq \frac{1}{\varepsilon}\left[v(\hat{x} + \varepsilon(y-\hat{x})) - v(\hat{x})\right]$$

$$\to \nabla v(\hat{x})^T (y-\hat{x})$$

as $\varepsilon \to 0$.

But $$\nabla v(\hat{x})^T = \left(\frac{\partial v}{\partial x_1}(\hat{x}),\ 0 \right)$$

So $$v(y) - v(\hat{x}) \geq \frac{\partial v}{\partial x_1}(\hat{x})(y_1 - 1).$$

But $\frac{\partial v}{\partial x_1}(\hat{x}) \le 0$, and for any $y \in F$

5 $y_1 - 1 \le 0$. So $v(y) \geq v(\hat{x})$, establishing that

$\hat{x}$ is a minimizer.

Optimization

1(b) Solution

2/12

$$v(y; u_0, u_1) = \tfrac{1}{2}(y-u_0)^2 + \tfrac{1}{2}(y-u_0-u_1)^2.$$

So

$$\frac{\partial v}{\partial u_0} = u_0 - y + u_0 + u_1 - y = 2u_0 + u_1 - 2y$$

$$\frac{\partial v}{\partial u_1} = u_0 + u_1 - y$$

Assume $|y| \le 1$

So $\nabla v = 0$ if $u_0 = y$, $u_1 = 0$.

4 As $\nabla_u^2 v(y; u_0, u_1) = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} > 0$.

v is convex

and $\hat{u}_0 = y$, $\hat{u}_1 = 0$ is a minimizer.

If $1 \le y \le 2$, take $\hat{u}_0 = 1$, $\hat{u}_1 = y - 1$

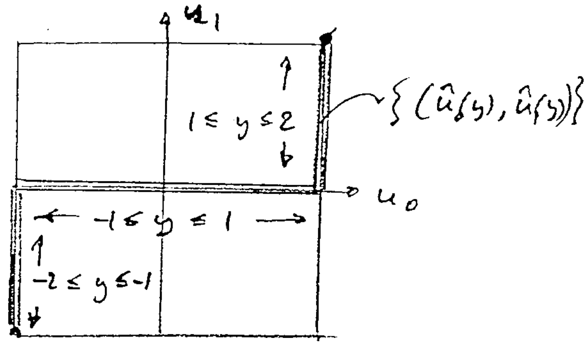
Then $\frac{\partial v}{\partial u_1}(y; 1, y-1) = 0$

and $\frac{\partial v}{\partial u_0}(y; 1, y-1) = 1 - y \le 0$.

4 So by part (a), $(1, y-1)$ is a minimizer of v over the square $\{|u_0| \le 1, |u_1| \le 1\}$. So $f(y) = 1$

is the desired feedback law on $1 \le y \le 2$.

2 The range of the minimizer pairs $(\hat{u}_0(y), \hat{u}_1(y))$



Optimization

2 Solution $\frac{7}{12}$

(a) $\hat{x}$ is an isolated local minimizer of v

if and only if

$$\nabla v(\hat{x}) = \underline{0} \quad ,$$

3

$$\nabla^2 v(\hat{x}) > 0 .$$

—

$$\nabla v_1(x_1, x_2) = \begin{bmatrix} x_2 + x_1^3 \\ x_1 + x_2^3 \end{bmatrix}$$

$$\nabla^2 v_1 = \begin{bmatrix} 3x_1^2 , & 1 \\ 1 , & 3x_2^2 \end{bmatrix}$$

4 (i) $\nabla v_1 = 0$ if $x = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, or if

$x_2 = -x_1^3 = x_2^9$; then $x = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ or $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$

$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ is not a local minimizer as $\nabla^2 v_1(0,0) =$

$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, which is not positive definite.

If $x = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ or $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ $\nabla^2 v_1 = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} > 0.$

So those are isolated local minimizers of v_1

—

(ii) $v_2 = 1 + (x_1 + x_2)^2 \geq 1.$

3 So $\begin{pmatrix} x_1 \\ -x_1 \end{pmatrix}$ is a local minimizer for all $x_1 \in \mathbb{R}$.

However $\begin{pmatrix} x_1 \\ -x_1 \end{pmatrix}$ is not an isolated local minimizer as $\begin{pmatrix} x_1 + \varepsilon \\ -x_1 - \varepsilon \end{pmatrix}$ is a local minimizer for all $\varepsilon > 0$.

Optimization

2(b) solution

Suppose x^c is the current iterate given by the method.

Calculate $\nabla v(x^c)$ and set $s = -\nabla v(x^c)$, a descent direction. The parameter w_1 in the next step $w_1 s$ is chosen as follows: choose $\mu < 1$ ($\mu = 0.8$, say).

6 Let $w_0 = \min \{ 2^k : k = \text{an integer}, \; v(x^c + 2^k s) \geq v(x^c) - \frac{2^k}{2} \|s\|^2 \}$

Let $w_1 = \max \{ \mu^n w_0 : n = 0, 1, 2, \ldots \; v(x^c + \mu^n w_0 s) < v(x^c) - \mu^n w_0 \|s\|^2 \}$

Then set the next iterate $x^+ = x^c + w_1 s$

Repeat until $\|x^+ - x^c\| <$ given error ε.

With careful choice of μ, the algorithm is decreasing; i.e. $v(x^+) < v(x^c)$. As the

4 level sets are bounded + $v \geq 0$ the sequence of iterates converges to one or other of the isolated local minimizers.

The convergence is at best linear.

Optimization

3 Solution

5/12

(a) $$\tilde{V}(x_n ; x_{n+1}) = V(x_n) + \nabla V(x_n)^T (x_{n+1} - x_n) + \frac{1}{2}(x_{n+1} - x_n)^T \nabla^2 V(x_n)(x_{n+1} - x_n)$$

But $$x_{n+1} - x_n = -\left(\nabla^2 V(x_n)\right)^{-1} \nabla V(x_n)$$

So
$$\tilde{V}(x_n : x_{n+1}) = V(x_n) - \nabla V(x_n)^T \nabla^2 V(x_n)^{-1} \nabla V(x_n) + \frac{1}{2} \nabla V(x_n)^T \nabla^2 V(x_n)^{-1} \nabla V(x_n)$$

8

$$= V(x_n) - \frac{1}{2} \nabla V(x_n)^T \nabla^2 V(x_n)^{-1} \nabla V(x_n)$$

$$< V(x_n)$$

the last inequality following from the positive-definiteness of the inverse Hessian and the fact that $\nabla V(x_n) \neq \underline{0}$.

Optimization
Solution

6/12

3(b) $V'(x) = 3(x^2-1)$, $V''(x) = 6x$

At $x=1$ $V'(x) = 0$, $V''(x) = 6$. So $x=1$

is a local minimizer. The Newton algorithm is

6 $$x_{n+1} = x_n - \frac{3(x_n^2-1)}{6x_n} = \frac{1}{2}(x_n + x_n^{-1})$$

If $x_0 = 1.1$. $x_1 = \frac{1}{2}(1.1 + .909) = 1.00454$

$x_2 = \frac{1}{2}(1.0045 + .995) = 1.00001$

(c) $s_0 = x_1 - x_0 = -.09545$

$s_1 = x_2 - x_1 = -.00453$

So since $\frac{s_{n+1}}{s_n^2} \approx \frac{s_n}{s_{n-1}^2}$ $s_2 \approx \frac{s_1^3}{s_0^2}$.

Then $s_2 \approx -1.02 \times 10^{-5}$

6

and $x_3 = 1.0000002$.

The convergence of the x_n to 1 would be quadratic — the number of zeroes in the decimal expansion of $x_n - 1$ is roughly doubling with each increase in n.

Optimization

7/2

4 Solution

(a) Suppose $\nabla v(\hat{x})^T z_i = 0$ for $i = 1, 2, \ldots, m$.

If $x \in x_0 + L$, $x \in \hat{x} + L$ + so for some $a_1, \ldots, a_m$.

8

$$x = \hat{x} + a_1 z_1 + \cdots + a_m z_m.$$

Hence

$$\nabla v(\hat{x})^T (x - \hat{x}) = \Sigma_1^m a_i \nabla v(\hat{x})^T z_i = 0$$

which, by the characterization given, implies that $\hat{x}$ is a minimizer.

(b) (i) Take $z_1 = (2, -1, 0)^T \quad (= (u_1, u_2, u_3)^T)$

Then $y_4 = (1, 2, 3) z_1 = 0$

Take $z_2 = (0 \ 3 \ -2)^T$: $y_4 = (1, 2, 3) z_2 = 0$.

5

z_1, z_2 are clearly linearly independent + so form a basis for the 2-dimensional L.

The constraint space is the linear variety $\frac{8}{12}$

$$x_0 + L = \{ (u_1, u_2, u_3)^T : u_1 + 2u_2 + 3u_3 = 1 \}$$

$$= (1,0,0)^T + L.$$

By (a), as $\nabla V(u_1,u_2,u_3) = \begin{pmatrix} 2u_1 \\ 2u_2 \\ 2u_3 \end{pmatrix}$

7 $\hat{u}$ is given by $\nabla V(\hat{u})^T z_1 = \nabla V(\hat{u})^T z_2 = 0$
and

$$\hat{u}_1 + 2\hat{u}_2 + 3\hat{u}_3 = 1.$$

That is,

$$2\hat{u}_1 - \hat{u}_2 = 0$$
$$3\hat{u}_2 - 2\hat{u}_3 = 0.$$

So $\hat{u}_1 = \frac{1}{2}\hat{u}_2$

$\hat{u}_3 = \frac{3}{2}\hat{u}_2$

$\left(\frac{1}{2} + 2 + \frac{9}{2}\right)\hat{u}_2 = 1$; so $\hat{u}_2 = \frac{1}{7}$.

So $(\hat{u}_1, \hat{u}_2, \hat{u}_3) = \left(\frac{1}{14}, \frac{1}{7}, \frac{3}{14}\right)$

Optimization

5 Solution

(a) The squared norm of the vector of linearized residuals is

$$\| r.(x;z)\|^2 = \Sigma_{k=1}^{m} \left(r_k(x) + \nabla r_k(x)^T(z-x)\right)^2.$$

Its (z)-gradient is

$$2\Sigma_{k=1}^{m} \left(r_k(x) + \nabla r_k(x)^T(z-x)\right)\nabla r_k(x)$$

7

As the squared norm is a positive definite quadratic expression in z, it is minimized if the gradient is zero:

$$\Sigma_{k=1}^{m} r_k(x)\nabla r_k(x) + \Sigma_{k=1}^{m} \nabla r_k(x)\nabla r_k(x)^T(z-x) = 0.$$

which is solved by $z = x^+$.

(b) Take the k^{th} residual to be the error term

$$r_k(x) = a + bp^k + cq^k - \bar{y}_k$$

and x to be $(a,b,c,p,q)^T$. The objective function

7 becomes

$$V(x) = \frac{1}{2}\Sigma_{k=1}^{100}\left(a + bp^k + cq^k - \bar{y}_k\right)^2$$

and $\hat{x} = (\hat{a},\hat{b},\hat{c},\hat{p},\hat{q})^T$ is the minimizer of this function. Its (x)-gradient is

$$\nabla V(x) = \Sigma_{k=1}^{100}\left(a + bp^k + cq^k - \bar{y}_k\right)\begin{bmatrix} 1 \\ p^k \\ q^k \\ kbp^{k-1} \\ kcq^{k-1} \end{bmatrix}$$

Optimization

5 solution

$\frac{10}{12}$

(c) Unlike the full Newton method, the Gauss-Newton method does not require the calculation at each step of the Hessian of second derivatives.

6 In "overdetermined" problems such as this ($100 > 5$) it works well if the residuals $r_k(\hat{x})$ are small or zero, as is assumed to be the case here. If the $r_k(\hat{x})$ are all zero it will converge quadratically. Otherwise it will converge at a fast linear rate.

Optimization

6 Solution

11/12

(a) s^+ minimizes $\frac{1}{2} s^T C s + b^T s$

subject to $s^T s = h^2$. So using the

Lagrange multiplier λ, it also minimizes

$$L(s,\lambda) = \frac{1}{2} s^T C s + b^T s + v(x^c) + \lambda(s^T s - h^2)$$

7 Hence. $\nabla_s L(s^+ \lambda^+) = 0$, $s^{+T} s^+ - h^2 = 0$

But $\nabla_s L = Cs + b + 2\lambda s$

So s^+ solves
$$c_{11} s_1 + c_{12} s_2 + b_1 + 2 s_1 \lambda = 0$$
$$c_{21} s_1 + c_{22} s_2 + b_2 + 2 s_2 \lambda = 0$$
$$s_1^2 + s_2^2 = h^2$$

— Eliminating λ gives the result.

(b) The function $v = x_1 x_2 + \frac{1}{4}(x_1^4 + x_2^4) + \frac{1}{3}(x_2^3 - x_1^3)$

has a gradient $\nabla v = \begin{bmatrix} x_2 + x_1^3 - x_1^2 \\ x_1 + x_2^3 + x_2^2 \end{bmatrix}$

and a Hessian $\nabla^2 v = \begin{bmatrix} 3x_1^2 - 2x_1 , & 1 \\ 1 , & 3x_2^2 + 2x_2 \end{bmatrix}$

So at $x^c = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $b = \nabla v(x^c) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

and

$$C = \nabla^2 v(x^c) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Solution of 6(b) continued

So a minimizer $s^+ = (s_1, s_2)^T$ solves

7 $$s_1^2 + s_2^2 = h^2$$

$$s_2^2 = s_1^2$$

– Possible solutions are four: $\left(\pm \frac{h}{\sqrt{2}}, \pm \frac{h}{\sqrt{2}}\right)$

(c): To determine the best choice: it is straightforward to evaluate $v(x^c + s^+)$ for each of the solutions and then choose the minimizer from the four solutions.

If $s_1 = s_2 = \frac{\pm h}{\sqrt{2}}$, $v(s_1, s_2) = \frac{h^2}{2} + \frac{h^4}{8}$

6 If $s_1 = -s_2 = \frac{h}{\sqrt{2}}$ $v(s_1, s_2) = -\frac{h^2}{2} - \frac{1}{3\sqrt{2}} h^3 + \frac{h^4}{8}$

If $s_1 = -s_2 = -\frac{h}{\sqrt{2}}$ $v(s_1, s_2) = -\frac{h^2}{2} + \frac{h^3}{3\sqrt{2}} + \frac{h^4}{8}$

So the minimizing step is $\left(\frac{h}{\sqrt{2}}, -\frac{h}{\sqrt{2}}\right)^T$

which coincides with x^+.