Some Examples, Questions, and Answers:
Lectures 6,7 – COMPSCI 369

Georgy Gimel'farb

Extremum Points

Example 6-7.1: The Taylor series decomposition of a multivariate function $f(\mathbf{x})$ at a stationary point $\mathbf{x} = \mathbf{a}$ where the gradient of $f(\mathbf{x})$ equal to zero:

$$\nabla f(\mathbf{a}) = \left[\begin{array}{c} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{array}\right]_{\mathbf{x}=\mathbf{a}} = \mathbf{0}$$

highlights the importance of the Hessian matrix of the second derivatives:

$$\mathbf{H}(\mathbf{a}) = \left[\begin{array}{cccc} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{array}\right]_{\mathbf{x}=\mathbf{a}}$$

for evaluating whether this point is a local minimum, a local maximum, or a saddle:

$$\begin{aligned} f(\mathbf{x}) &= f(\mathbf{a}) + (\mathbf{x}-\mathbf{a})^{\mathsf{T}} \nabla f(\mathbf{a}) + \tfrac{1}{2}(\mathbf{x}-\mathbf{a})^{\mathsf{T}} \mathbf{H}(\mathbf{a})(\mathbf{x}-\mathbf{a}) + O(|\mathbf{x}-\mathbf{a}|^3) \\ &\approx f(\mathbf{a}) + \tfrac{1}{2}(\mathbf{x}-\mathbf{a})^{\mathsf{T}} \mathbf{H}(\mathbf{a})(\mathbf{x}-\mathbf{a}) \end{aligned}$$

In other words, in the vicinity of the stationary point $\mathbf{a}$ where the absolute deviation from this point (i.e. the length, or norm of the vector $|\mathbf{x}-\mathbf{a}|$ is very small), the behaviour of the function is specified by its close approximation with the quadratic form $f(\mathbf{a}) + \frac{1}{2}(\mathbf{x}-\mathbf{a})^{\mathsf{T}} \mathbf{H}(\mathbf{a})(\mathbf{x}-\mathbf{a})$.

Let us consider the bivariate function $f(x,y) = x^2 + y^2(1-x)^3$ with a single stationary point $x = y = 0$ (see `http://en.wikipedia.org/wiki/Local_extremum`):

$$\frac{\partial f(x,y)}{\partial x} = 2x - 3y^2(1-x)^2; \quad \frac{\partial f(x,y)}{\partial y} = 2y(1-x)^3;$$

$$\frac{\partial^2 f(x,y)}{\partial x^2} = 2 - 6y^2(1-x); \quad \frac{\partial^2 f(x,y)}{\partial x \partial y} = -6y(1-x)^2; \quad \frac{\partial^2 f(x,y)}{\partial^2 y} = 2(1-x)^2$$

There are no other stationary points because for the system of equations $\nabla f(\mathbf{x}) = \mathbf{0}$:

$$\left\{ \begin{array}{rcl} 2x - 3y^2(1-x)^2 & = & 0 \\ 2y(1-x)^3 & = & 0 \end{array} \right.$$

$y = 0$ implies $x = 0$ and $x = 0$ implies $y = 0$, but for any $y \neq 0$ both the equations cannot be set to zero simultaneously: the second gradient component is equal to zero for $x = 1$, and this value does not match the first equation.

The Hessian $\mathbf{H}(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ is positive definite, so $f(0,0) = 0$ is a local minimum.

Note that this local minimum says nothing about the global behaviour of this function. Actually, its range spans from ∞ for $x > 1$ and $y \to \pm\infty$ to $+\infty$ for $x < 1$ and $y \to \pm\infty$.

Question 6-7.1: Specify stationary points of the function $f(x, y, z) = x^2 + y^2 + z^2$.

Question 6-7.2: Specify stationary points of the function $f(x, y, z) = x^2 + y^2 - z^2$.

Example 6-7.2: (see `http://en.wikipedia.org/wiki/Saddle_point`) The Sylvester's criterion gives a sufficient condition of whether the point is a local maximum, minimum, or saddle. If the Hessian matrix at this point is the null matrix, then the quadratic term in the Taylor's decomposition in Example 6-7.1 is absent, and the behaviour of $f(\mathbf{x})$ is specified by the higher-order derivatives.

Let us consider the function $f(x, y, z) = x^4 + y^4 + z^4$ with the gradient $\nabla f(x, y, z) = [4x^3, 4y^3, 4z^3]^\mathsf{T}$ and Hessian $\mathbf{H}(x, y, z) = \mathrm{diag}\{12x^2, 12y^2, 12z^2\}$. The point $(0, 0, 0)$ is the local minimum, but the Hessian $\mathbf{H}(0, 0, 0) = \mathbf{0}$ is the null matrix, which is neither definite, nor indefinite.

Example 6-7.3: Specifying stationary points of the bivariate Rosenbrock's "Banana" function $f(x, y) = (1 - x)^2 + 100(y - x^2)^2$ (the colour-coded graphs of this function are shown in Fig. 1; see also Lecture Notes `CS-Optimisation-1.pdf`, Slide 17):

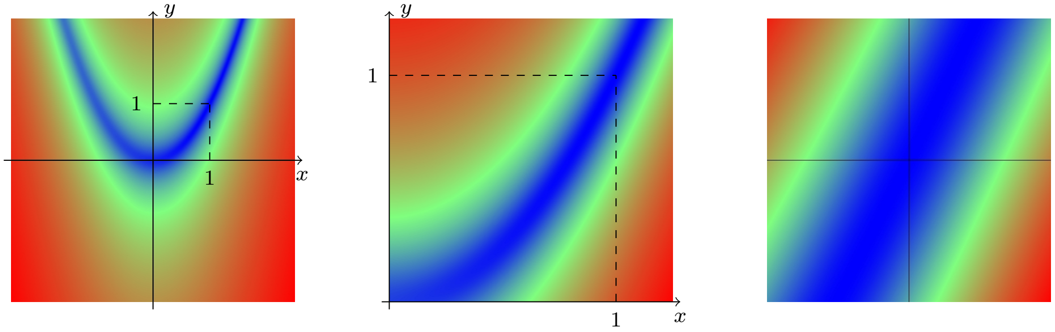


Figure 1: Logarithmic colour scales are used for visualisation because of the fast growth of function values across the walls of the "banana valley"; the minimum is at $f(1, 1) = 0$.

Left window $(-2.5 \leq x, y \leq 2.5)$: $f(-2.5, -2.5) = 7668.5$; $f(0, 0) = 1$;

Central window $(0 \leq x, y \leq 1.25)$: $f(0, 0) = 1$; $f(1.25, 0) = 244.2$

Right window $(0.875 \leq x, y \leq 1.125)$: $f(0.875, 0.875) \approx 1.21$; $f(1.061, 1.125) \approx 0.37$

Zero gradient $\nabla f = \begin{bmatrix} -2(1-x) - 400x(y-x^2) \\ 200(y-x^2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ implies the single stationary point $(x = 1, y = 1)$ of this function: $x = y^2$ from the second equation and then $x = 1$ from the first equation.

The Hessian matrix $\mathbf{H}(x, y) = \begin{bmatrix} 2 - 400y + 1200x^2 & -400 \\ -400 & 200 \end{bmatrix}$ is positive definite at the point $(1, 1)$ due to the positive determinants for the upper-left sunmatrices (namely, 802 and $802 \cdot 200 - 400^2 = 400$, so this point is the local minimum.

Note that values of this function form a very narrow valley with steep walls, such that its bottom follows the quadratic line $y = x^2$. The bottom-line values are growing also quadratically, but much slower than across the walls: $f(x, y = x^2) = (1 - x)^2$, around the minimum point $(0, 0)$, so that $f(0, 0) = f(2, 4) = 1$, whereas $f(1.1, 0) = 146.42$. As Fig. 2 suggests, this pattern of changing may hinder the line search for the goal minimum along the gradient direction.

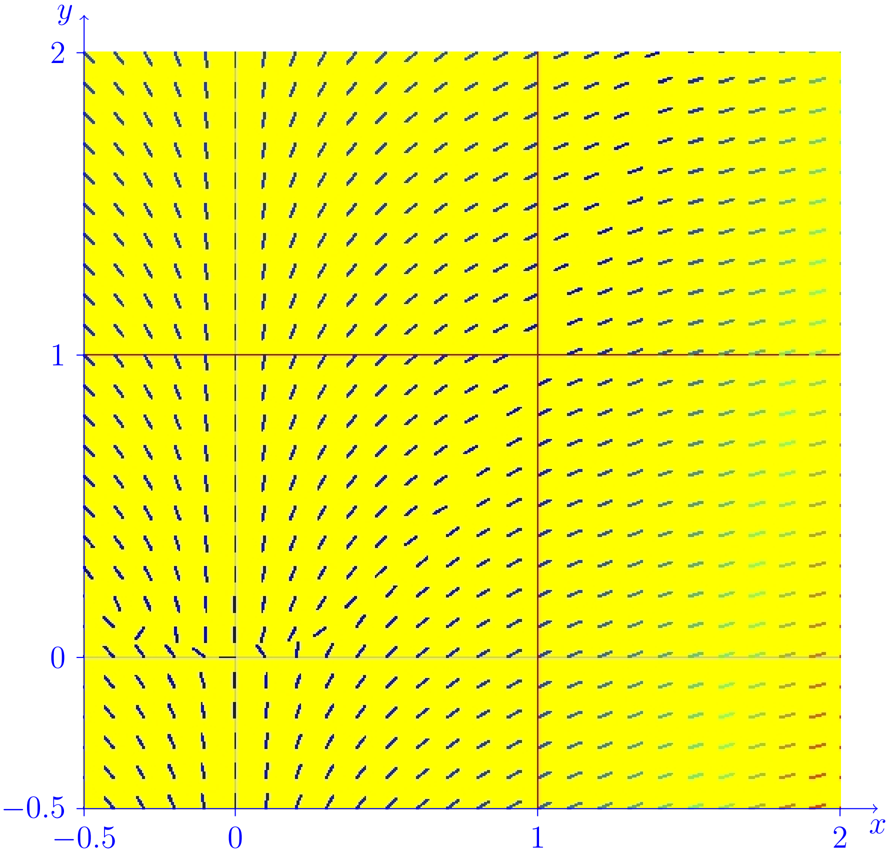


Figure 2: Needle gradient map of the Rosenbrock's "Banana" function: $-0.5 \leq x \leq 2$; $-0.5 \leq y \leq 2$; step of 0.1 along both x- and y-axes; colour coding of gradient magnitude within the range $[0, 3712.735]$ (the minimum and maximum magnitude at the point $(1, 1)$ and $(2, -0.5)$, respectively), which is mapped linearly onto the colour scale from blue to red.

Line Search

Example 6-7.4: Searching for a local minimum of $\varphi(x,y) = 2x^3 + y^3 - 6x - 3y$ (the cubic function).

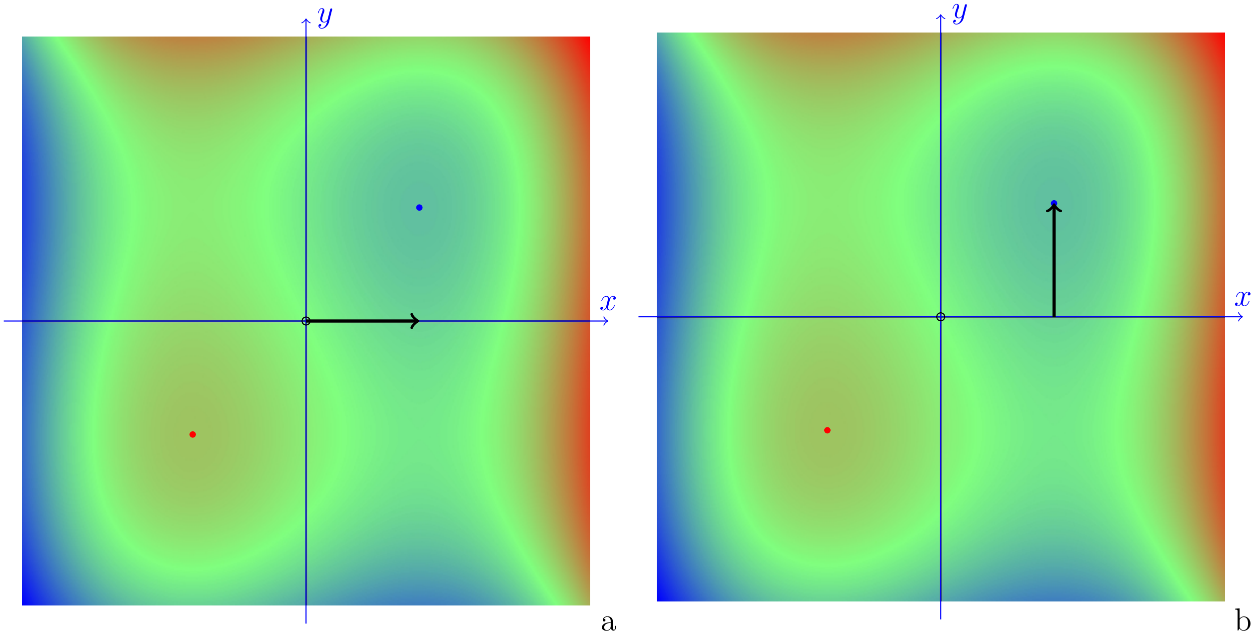


Figure 3: Colour-coded graph of the cubic function $\varphi(x,y)$ in the range $-2.5 \leq x, y \leq 2.5$. Function values are between $\varphi(-2.5,-2.5) = -24.375$ and $\varphi(2.5,2.5) = 24.375$. The function has four singular points with zero gradient $\nabla\varphi(x,y) = \left[6x^2 - 6,\ \ 3y^2 - 3\right]^{\mathsf{T}}$, namely, $(x = \pm 1, y = \pm 1)$. According to the Hessian diag $\{12x, 6y\}$ $\varphi(1,1) = -6$ is at the local minimum; $\varphi(-1,-1) = 6$ is at the local maximum, and the $\varphi(-1,1) = 2$ and $\varphi(1,-1) = -2$ are at the saddle points.

- **Search from** $(0,0)$ **along the line** $x(t) = t, y(t) = 0$ (see Fig. 3,a).

  Substitution of the line coordinates into the function $f(x,y)$ of two variables produces the function of one variable $\varphi(t) = 2t^3 - 6t$.

  Its first derivative, $\varphi'(t) = 6t^2 - 6$, suggests that the function has two stationary points: $\varphi'(x) = 0$ for $t = \pm 1$.

  The values of the second derivative, $\varphi''(t) = 12t$, at these points show that the local minimum is in the point $t = 1$: $\varphi(1) \equiv f(1,0) = -4$ and the local maximum is in the point $t = -1$: $\varphi(-1) \equiv f(-1,0) = 4$. Therefore, this line search gives the minimiser $t^* = 1$, i.e. the point $(x^* = 1, y^* = 0)$.
- **The like search from** $(1,0)$ **along the line** $x(s) = 1, y(s) = s$ results in the univariate function $\psi(s) = s^3 - 3s - 4$ with the stationary points for $s = \pm 1$, i.e. $(x = 1, y = 1)$ and $(x = 1, y = -1)$, and the minimiser $s^* = 1$, i.e. the point $x^* = 1, y^* = 1$ (see Fig. 3,b).
- The above line searches reached the actual local minimum due to the right, by pure chance, selection of the successive line directions.

Question 6-7.3: What point will the second search from $(1,0)$ reach, be it performed along the line $x(s)=1+s, y(s)=s$?

Example 6-7.5: Maximising an univariate function $\varphi(t)$, which is provably unimodal, i.e. has exactly one maximum, within a given interval $[a,b]$.

The maximiser $t^* = \arg\max_t \varphi(t)$; $a \le t^* \le b$, separates the interval $[a,b]$ into the two parts: the function $\varphi(t)$ is increasing monotonously within $[a,t^*]$ and decreasing monotonously within $[t^*,b]$:

$$\begin{array}{llll} \text{if} & a < t < t^* & \text{then} & \varphi(a) < \varphi(t) < \varphi(t^*) \\ \text{if} & t^* < t < b & \text{then} & \varphi(t^*) > \varphi(t) > \varphi(b) \end{array}$$

Let $a < t_1 < t_2 < b$. Then $\varphi(t_1) > \varphi(a)$ and $\varphi(t_2) > \varphi(b)$, and the interval for t^* can be reduced, depending on whether $\varphi(t_1)$ is lesser than, equal to, or greater than $\varphi(t_2)$.

$\varphi(t_1) < \varphi(t_2)$: Then $\varphi(a) < \varphi(t_1) < \varphi(t_2) \le \varphi(t^*)$, so $[a,b]$ can be reduced to $[t_1,b]$.

$\varphi(t_1) = \varphi(t_2)$: Then $\varphi(a) < \varphi(t_1) = \varphi(t_2) \le \varphi(t^*)$, so $[a,b]$ can be reduced to $[t_1,t_2]$.

$\varphi(t_1) > \varphi(t_2)$: Then $\varphi(b) < \varphi(t_2) < \varphi(t_1) \le \varphi(t^*)$, so $[a,b]$ can be reduced to $[a,t_2]$.

Golden section search follows this simple scheme, providing an initial interval $a,b]$, which contains the maximiser t^* of a unimodal function $\varphi(t)$ is known:

```
Require: interval bounds $a$; $b$ $(a < b)$; a function $\varphi(t)$; a precision threshold $\varepsilon$
  $\gamma_b = 0.5(\sqrt{5}-1) \approx 0.618$;    $\gamma_a = 1 - \gamma_b$;
  while $b - a > \varepsilon$ do
    $t_1 = a + \gamma_a \cdot (b-a)$;    $t_2 = a + \gamma_b \cdot (b-a)$;
    if $u(t_1) < u(t_2)$ then $a = t_1$;
    else
      if $u(t_1) = u(t_2)$ then $a = t_1$; $b = t_2$;
      else $b = t_2$ ;
      end if
    end if
  end while
```

Question 6-7.4: Describe the golden section search for the minimiser of a unimodal function in the interval $[a,b]$ that contains the goal minimiser.

Gradient Search

Example 6-7.6: Gradient search for the minimiser of the quadratic function $f(x,y) = ax^2 + bxy + cy^2 + dx + ey$

Require: the function $f(x,y)$; its gradient

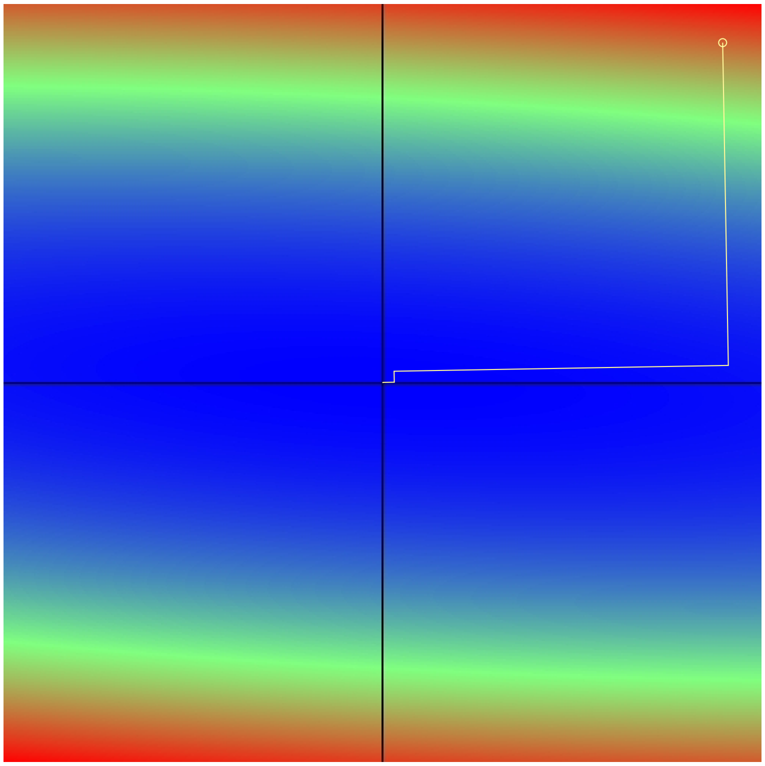
$$\nabla f(x,y) = \left[\begin{array}{l} \nabla_x(x,y) = 2ax + by + d \\ \nabla_y(x,y) = bx + 2cy + e \end{array} \right];$$

a starting point (x_0,y_0); a precision threshold ε

```
$x = x_0$; $y = y_0$;
while $|\varepsilon_x| > \varepsilon$ AND $|\varepsilon_y| > \varepsilon$ AND $|\varepsilon_f| > \varepsilon$ do
    $u = \nabla_x(x, y)$; $v = \nabla_y(x, y)$;
    $t_n = 2axu + b(xv + yu) + 2cyv + du + ev$; $t_d = 2(au^2 + buv + cv^2)$;
    $t^\circ = t_n / t_d$;
    $x^\circ = x - t^\circ u$; $y^\circ = y - t^\circ v$;
    $\varepsilon_x = x - x^\circ$; $\varepsilon_y = y - y^\circ$; $\varepsilon_f = f(x, y) - f(x^\circ, y^\circ)$;
    $x = x^\circ$; $y = y^\circ$;
end while
```

For $f(x, y) = x^2 - 3xy + 30y^2$, $(x_0 = 9.0, y_0 = 9.0)$, and $\varepsilon = 0.0001$:

Step	x	y	$f(x,y)$	$\nabla_x(x,y)$	$\nabla_y(x,y)$	x°	y°
0	9.0000	9.0000	2268.0000	-9.0000	513.0000	9.1498	0.4624
1	9.1498	0.4624	77.4404	16.9123	0.2967	0.3073	0.3073
2	0.3073	0.3073	2.6442	-0.3073	17.5163	0.3124	0.0158
3	0.3124	0.0158	0.0903	0.5775	0.0101	0.0105	0.0105
4	0.0105	0.0105	0.0031	-0.0105	0.5981	0.0107	0.0005
5	0.0107	0.0005	0.0001	0.0197	0.0003]	0.0004	0.0004
6	0.0004	0.0004	0.0000	-0.0004	0.0204]	0.0004	0.0000
7	0.0004	0.0000	0.0000	0.0007	0.0000]	0.0000	0.0000
8	0.0000	0.0000	0.0000	0.0000	0.0007]	0.0000	0.0000
9	0.0000	0.0000	0.0000	-0.0105	0.2003]	0.0000	0.0000



Question 6-7.5: How will this search proceed from the starting point $(5.0, 20.0)$?

Answers

To Question 6-7.1: The gradient $\nabla f(x,y,z) = [2x, 2y, 2z]^\mathsf{T}$ has one stationary point at $(x = 0, y = 0, z = 0)$. Because the Hessian in this case is the positive definite diagonal matrix $\mathbf{H} = \mathrm{diag}\{2,2,2\}$ (due to the regular sequence $\{2,4,8\}$ of the positive determinants of its upper-left submatrices), it is a local minimum with $f(0,0,0) = 0$.

To Question 6-7.2: The gradient $\nabla f(x,y,z) = [2x, 2y, -2z]^\mathsf{T}$ has one stationary point at $(x = 0, y = 0, z = 0)$. The Hessian in this case is the indefinite diagonal matrix $\mathbf{H} = \mathrm{diag}\{2,2,-2\}$ (due to the irregular sequence $2, 4, -8$ of the determinants of its upper-left submatrices). Thus, it is a saddle point with $f(0,0,0) = 0$.

To Question 6-7.3: In this case the univariate function $\psi(s) = 2(1+s)^3 + s^3 - 6(1+s) - 3s = 3s^3 + 6s^2 - 3s - 4$ and its first derivative $\psi'(s) = 9s^2 + 12s - 3$ suggests two stationary points $\psi'(s^a st) = 0$, namely,

$$s_1^* = \frac{1}{3}\left(-2 + \sqrt{7}\right) \approx 0.216 \text{ and } s_1^* = \frac{1}{3}\left(-2 - \sqrt{7}\right) \approx -1.560$$

The second derivative $\psi''(s) = 18s + 12 = \pm 6\sqrt{7}$ shows that the first point is the local minimum and the second point is the local maximum. Therefore, the minimiser is $s_1^* \approx 0.216$, and the search reaches the point $\left(x^* = \frac{1}{3}\left(1 + \sqrt{7}\right) \approx 1.216, y^* = \frac{1}{3}\left(-2 + \sqrt{7}\right) \approx 0.216\right)$.

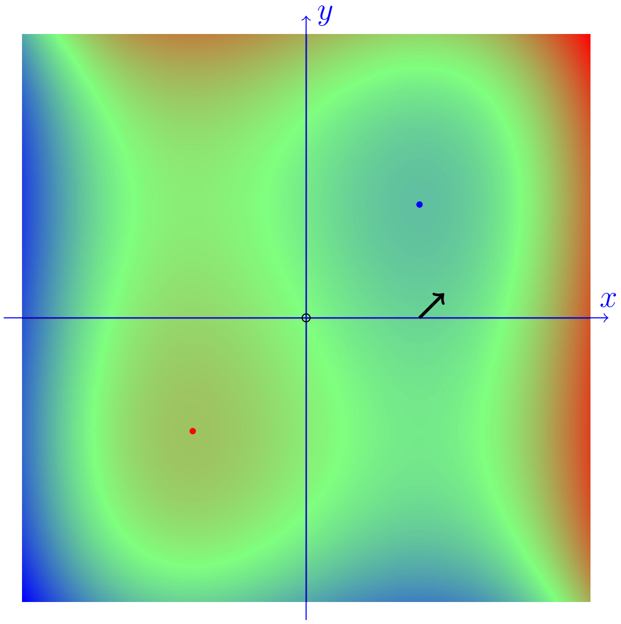


Figure 4: Colour-coded graph of the cubic function $\varphi(x,y)$ as in Fig. 3: the gradient search for the minimum along the line $x(s) = 1 + s, y(s) = s$. The minimiser $s^* \approx 0.216$, i.e. the minimum along the line is in the point $(1.216, 0.216)$.

To Question 6-7.4: The goal minimiser, $t^* = \arg\min_t \varphi(t)$; $a \leq t^* \leq b$, separates the interval $[a, b]$ into two parts: of the monotone decrease of $\varphi(t)$ within $[a, t^*]$ and monotone increase within $[t^*, b]$:

$$\begin{array}{llll} \text{if} & a < t < t^* & \text{then} & \varphi(a) > \varphi(t) > \varphi(t^*) \\ \text{if} & t^* < t < b & \text{then} & \varphi(t^*) < \varphi(t) < \varphi(b) \end{array}$$

Let $a < t_1 < t_2 < b$. Then $\varphi(t_1) < \varphi(a)$ and $\varphi(t_2) < \varphi(b)$, and the interval for t^* can be reduced, depending on whether $\varphi(t_1)$ is lesser than, equal to, or greater than $\varphi(t_2)$.

$\varphi(t_1) < \varphi(t_2)$: Then $\varphi(b) > \varphi(t_2) > \varphi(t_1) \leq \varphi(t^*)$, so $[a, b]$ can be reduced to $[a, t_2]$.

$\varphi(t_1) = \varphi(t_2)$: Then $\varphi(a) < \varphi(t_1) = \varphi(t_2) \leq \varphi(t^*)$, so $[a, b]$ can be reduced to $[t_1, t_2]$.

$\varphi(t_1) > \varphi(t_2)$: Then $\varphi(a) > \varphi(t_1) > \varphi(t_2) \leq \varphi(t^*)$, so $[a, b]$ can be reduced to $[t_1, b]$.

Golden section search for the minimiser is then as follows:

```
Require: interval bounds $a$; $b$ ($a < b$); a function $\varphi(t)$; a precision threshold $\varepsilon$
  $\gamma_b = 0.5(\sqrt{5} - 1) \approx 0.618$;    $\gamma_a = 1 - \gamma_b$;
  while $b - a > \varepsilon$ do
      $t_1 = a + \gamma_a \cdot (b - a)$;    $t_2 = a + \gamma_b \cdot (b - a)$;
      if $u(t_1) < u(t_2)$ then $b = t_2$;
      else
          if $u(t_1) = u(t_2)$ then $a = t_1$; $b = t_2$;
          else $a = t_1$ ;
          end if
      end if
  end while
```

To Question 6-7.5:

Step	x	y	$f(x, y)$	$\nabla_x(x, y)$	$\nabla_y(x, y)$	x°	y°
0	5.0000	20.0000	11725.0000	−50.0000	1185.0000	5.8313	0.2291
1	5.8313	0.2291	31.4550	10.7651	0.4542	0.0134	0.0537
2	0.0134	0.0537	0.0844	−0.1341	3.1790	0.0156	0.0008
3	0.0156	0.0008	0.0002	0.0289	0.0012	0.0000	0.0001
4	0.0000	0.0001	0.0000	−0.0004	0.0085	0.0000	0.0000
5	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000