Optimization

Class 4: Practical Optimization

1 Basic Optimization Method

We've seen that algebra is not practical for finding stationary points (and hence optimizers) of general functions; now we'll introduce an approach that is practical.

Basic Optimization Method

1. **Initialize** the current guess;
2. **Update**: Determine a new guess from the current guess;
3. **Stop or repeat**: If stopping criterion is not satisfied and progress is being made, use the new guess as the current guess and repeat from step 2.

KEY QUESTION: How to update intelligently?

2 Traditional Approach

Replace the actual objective function f with a simpler function that "matches" f near the current guess, and whose stationary point is easily found; then use this stationary point as the next guess. Linear functions very simple, but they rarely have stationary points, and when they do, every point in the input space is stationary (e.g., their graph is a plane parallel to the xy-plane). A slightly more complicated function whose stationary point usually exists (and can be found easily) is a quadratic.

Q: For what kind of functions f could there be a matching quadratic everywhere?

A: When f is quadratic to begin with, we can "replace" it with itself to match everywhere.

2.1 Quadratics with two variables (x, y)

$$Q(x,y) = \frac{a}{2}\,x^2 + b\,x\,y + \frac{c}{2}\,y^2 + d\,x + g\,y + h$$

for constants a, b, c, d, g, and h. The gradient (vector) of Q is

$$\nabla Q(x,y) = \begin{bmatrix} a\,x + b\,y + d \\ b\,x + c\,y + g \end{bmatrix},$$

so the stationary points of Q are the solutions (x, y) to

$$\begin{aligned} a\,x + b\,y + d &= 0 \\ b\,x + c\,y + g &= 0. \end{aligned} \tag{1}$$

Group Problem 1

(a) Solve the pair of equations (1).

(b) Choose values of the constants for which there are multiple solutions.

(c) Choose values of the constants for which there is no solution.

Group Problem 1 (solution)

(a)

$$x = \frac{-b\,g + c\,d}{b^2 - a\,c} \qquad y = \frac{a\,g - b\,d}{b^2 - a\,c}.$$

(b) There are infinitely many solutions if $b^2 = a\,c$, $c\,d = b\,g$, and $a\,g = b\,d$; since then we can multiply the top equation by c (or b) and the bottom by $-b$ (or $-a$) and add to get $c\,d = b\,g$ (or $a\,g = b\,d$).

(c) There is no solution if $b^2 = a\,c$ and either $c\,d \neq b\,g$ or $a\,g \neq b\,d$ since then multiplying the top equation by c (or b) and the bottom equation by $-b$ (or $-a$) and adding gives a contradiction $c\,d = b\,g$ (or $a\,g = b\,d$).

It should not be a surprise that there are quadratic functions for which no stationary point exists, since linear functions often have that property and are included in the category of quadratic functions (with $a = b = c = 0$).

Group Problem 2

(a) Compute $Q_x(0,0)$, $Q_y(0,0)$, Q_{xx}, Q_{xy}, and Q_{yy} for the quadratic

$$Q(x,y) = \frac{a}{2}\,x^2 + b\,x\,y + \frac{c}{2}\,y^2 + d\,x + g\,y + h.$$

(b) Use matrix notation to express $Q(x,y)$ in terms of $Q(0,0)$, the vector $\begin{bmatrix} x \\ y \end{bmatrix}$, the matrix $\begin{bmatrix} Q_{xx} & Q_{xy} \\ Q_{xy} & Q_{yy} \end{bmatrix}$, and the gradient vector $\nabla Q(0,0) = \begin{bmatrix} Q_x(0,0) \\ Q_y(0,0) \end{bmatrix}$.

Group Problem 2 (solution)

(a) $Q_x(0,0) = d$, $Q_y(0,0) = g$, $Q_{xx} = a$, $Q_{xy} = b$, $Q_{yy} = c$.

(b)

$$\begin{aligned} Q(x,y) &= \frac{1}{2}\begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} d \\ g \end{bmatrix}^T \begin{bmatrix} x \\ y \end{bmatrix} + h \\ &= \frac{1}{2}\begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} Q_{xx} & Q_{xy} \\ Q_{xy} & Q_{yy} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \nabla Q(0,0)^T \begin{bmatrix} x \\ y \end{bmatrix} + Q(0,0). \end{aligned}$$

Note that the superscript T indicates "transpose", which turns a column vector into a row vector for dot-product multiplication:

$$\text{e.g.} \quad \begin{bmatrix} d \\ g \end{bmatrix}^T \begin{bmatrix} x \\ y \end{bmatrix} = [d \quad g] \begin{bmatrix} x \\ y \end{bmatrix} = d\,x + g\,y.$$

Now we'll explore how to choose the coefficients a, b, c, d, g, and h to make the quadratic match f near a given base point.

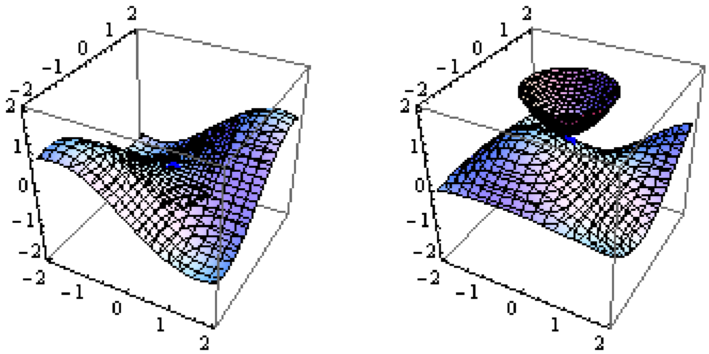


Figure 1: A quadratic that matches, and one that doesn't

3 Matching $f(x, y)$ with the Taylor Quadratic

When the base point is $(0, 0)$, a good choice is the quadratic

$$\frac{1}{2} \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} f_{xx}(0,0) & f_{xy}(0,0) \\ f_{xy}(0,0) & f_{yy}(0,0) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \nabla f(0,0)^T \begin{bmatrix} x \\ y \end{bmatrix} + f(0,0),$$

which matches (the value as well as the first and second derivatives of) f at $(0, 0)$.

For a general base point $(\bar{x}, \bar{y})$, the *Taylor quadratic* TQ matches the value and derivatives of f at the base point $(\bar{x}, \bar{y})$:

$$\boxed{TQ(x,y) := \frac{1}{2} \begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix}^T \nabla^2 f(\bar{x},\bar{y}) \begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix} + \nabla f(\bar{x},\bar{y})^T \begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix} + f(\bar{x},\bar{y})}$$

in terms of the *second-derivative matrix*

$$\nabla^2 f(\bar{x},\bar{y}) = \begin{bmatrix} f_{xx}(\bar{x},\bar{y}) & f_{xy}(\bar{x},\bar{y}) \\ f_{xy}(\bar{x},\bar{y}) & f_{yy}(\bar{x},\bar{y}) \end{bmatrix}.$$

See Contours of Taylor quadratic in Class4.nb

3.1 Discretization example

Consider the integral optimization problem:

$$\min \int_0^1 x(t)^2\, x'(t)\, \mathrm{d}t \text{ with } x(0) = 0 \text{ and } x(1) = 1$$

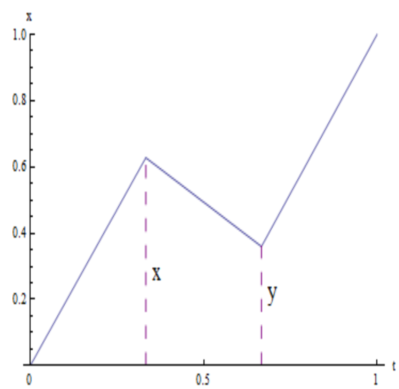
Group Problem 3

(a) Find the objective function $f(x, y)$ in the right-endpoint discretization at the t-values 1/3 and 2/3.

(b) Compute the Taylor quadratic $TQ(x, y)$ at base point $(\bar{x}, \bar{y}) = (1, 2)$ for this $f(x, y)$, convert it to non-matrix form, and confirm that $\nabla TQ(x, y) = \nabla^2 f(\bar{x}, \bar{y}) \begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix} + \nabla f(\bar{x}, \bar{y})$ in this case.

Group Problem 3 (solution)

(a)



We label x as the height of the right endpoint of the discretized function on the first subinterval $\left[0, \frac{1}{3}\right]$, and use the substitutions $x(t) \to x$ and $x'(t) \to \dfrac{x}{\frac{1}{3}}$ to convert that part of the integral:

$$\int_0^{\frac{1}{3}} x(t)^2\, x'(t)\, \mathrm{d}t \to \int_0^{\frac{1}{3}} x^2 \left(\frac{x}{\frac{1}{3}}\right) \mathrm{d}t \to x^2 \left(\frac{x}{\frac{1}{3}}\right)\left(\frac{1}{3}\right) = x^3.$$

We label y as the height of the right endpoint of the discretized function on the second subinterval $\left[\frac{1}{3}, \frac{1}{3}\right]$, and use the substitutions $x(t) \to y$ and $x'(t) \to \dfrac{y - x}{\frac{1}{3}}$ to convert that part of the integral:

$$\int_0^{\frac{1}{3}} x(t)^2\, x'(t)\, \mathrm{d}t \to \int_0^{\frac{1}{3}} y^2 \left(\frac{y - x}{\frac{1}{3}}\right) \mathrm{d}t \to y^2 \left(\frac{y - x}{\frac{1}{3}}\right)\left(\frac{1}{3}\right) = y^3 - x\, y^2.$$

We know the height of the right endpoint of the discretized function is 1 on the final subinterval $\left[\frac{2}{3}, 1\right]$, so we use the substitutions $x(t) \to 1$ and $x'(t) \to \dfrac{1-y}{\frac{1}{3}}$ to convert that part of the integral:

$$\int_0^{\frac{1}{3}} x(t)^2\, x'(t)\,\mathrm{d}t \to \int_0^{\frac{1}{3}} 1^2 \left(\frac{1-y}{\frac{1}{3}}\right) \mathrm{d}t \to 1^2 \left(\frac{1-y}{\frac{1}{3}}\right)\left(\frac{1}{3}\right) = 1-y.$$

Adding the three subinterval approximations together gives us the objective function we seek: $\boxed{f(x,y) = x^3 - x\,y^2 + y^3 - y + 1}$.

(b)

$$\begin{aligned}
f_x(x,y) &= 3\,x^2 - y^2\\
f_y(x,y) &= -2\,x\,y + 3\,y^2 - 1\\
f_{xy}(x,y) &= -2\,y\\
f_{xx}(x,y) &= 6\,x\\
f_{yy}(x,y) &= -2\,x + 6\,y.
\end{aligned}$$

$$\begin{aligned}
f(1,2) &= 4\\
f_x(1,2) &= -1\\
f_y(1,2) &= 7\\
f_{xy}(1,2) &= -4\\
f_{xx}(1,2) &= 6\\
f_{yy}(1,2) &= 10,
\end{aligned}$$

$$\begin{aligned}
TQ(x,y) &= \frac{1}{2}\begin{bmatrix} x-1 \\ y-2 \end{bmatrix}^T \begin{bmatrix} 6 & -4 \\ -4 & 10 \end{bmatrix}\begin{bmatrix} x-1 \\ y-2 \end{bmatrix} + \begin{bmatrix} -1 \\ 7 \end{bmatrix}^T \begin{bmatrix} x-1 \\ y-2 \end{bmatrix} + 4\\
&= 3(x-1)^2 - 4(x-1)(y-2) + 5(y-2)^2 - 1(x-1) + 7(y-2) + 4.
\end{aligned}$$

From this, we immediately get

$$\nabla T(x,y) = \begin{bmatrix} 6(x-1) - 4(y-2) - 1 \\ -4(x-1) + 10(y-2) + 7 \end{bmatrix}$$

by computing the partial derivatives of T with respect to x and y respectively. This result is evidently the same as

$$\nabla^2 f(\bar{x},\bar{y}) \begin{bmatrix} x-\bar{x} \\ y-\bar{y} \end{bmatrix} + \nabla f(\bar{x},\bar{y}) = \begin{bmatrix} 6 & -4 \\ -4 & 10 \end{bmatrix}\begin{bmatrix} x-1 \\ y-2 \end{bmatrix} + \begin{bmatrix} -1 \\ 7 \end{bmatrix}.$$

There is a nice analogue in general here to the differentiation of a single-variable quadratic $Q(x) = \frac{1}{2}\,A\,x^2 + B\,x + C$ with coefficients A, B, and C. In that case we can compute $Q'(x) = A\,x + B$. Notice that the squared-term in the original quadratic can be rewritten so that the quadratic can be expressed as follows $Q(x) = \frac{1}{2}\,x\,A\,x + B\,x + C$; which is exactly the pattern in our two-variable Taylor quadratic but where $\nabla^2 f(\bar{x}, \bar{y})$ plays the role of A and $\nabla f(\bar{x}, \bar{y})$ plays the role of B. These same roles are found in the comparison of $Q'(x) = A\,x + B$ with $\nabla TQ(x, y) = \nabla^2 f(\bar{x}, \bar{y}) \begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix} + \nabla f(\bar{x}, \bar{y})$.