1 Optimization

Reading [Simon], Chapter 17, p. 396-410.

1.1 Recall One Variable Case

Let $f : U \to R$ be a one variable and real valued function whose domain is an open set $U \subset R$.

A point $x^* \in U$ is a *critical point* for f if $f'(x^*) = 0$. This condition is necessary condition for *local extremum*:

If x^* is a point of local maximum or minimum then $f'(x^*) = 0$.

But of course this is not a sufficient condition, see for example $f(x) = x^3$.

The sufficient condition is not vanishing of second derivative:

If $f'(x^*) = 0$ and $f''(x^*) < 0$ ($f''(x^*) > 0$) then x^* is a local maximum (minimum) point.

1.2 Multivariable Case

Let $F : U \to R$ be an n-variable and real valued function whose domain is an open set $U \subset R^n$.

First some definitions.

1. A point $x^* = (x_1^*, \ ... \ , x_n^*) \in U$ is a **global max** of f if
$$F(x^*) \geq F(x) \quad for \quad all \quad x \in U.$$

2. A point $x^* = (x_1^*, \ ... \ , x_n^*) \in U$ is a **local max** of f if
$$F(x^*) \geq F(x) \quad for \quad all \quad x \in U \bigcap B_r(x^*),$$

for *some* ball $B_r(x^*)$.

3. A point $x^* = (x_1^*, \ ... \ , x_n^*) \in U$ is a **strict global max** of f if
$$F(x^*) > F(x) \quad for \quad all \quad x \in U.$$

4. A point $x^* = (x_1^*, \ ... \ , x_n^*) \in U$ is a **strict local max** of f if
$$F(x^*) > F(x) \quad for \quad all \quad x \in U \bigcap B_r(x^*),$$

for *some* ball $B_r(x^*)$.

Similarly are defined **min**-s.

Examples

1. $F(x_1, x_2) = x_1^2 + x_2^2$ has strict global minimum at $x^* = (0, 0)$, check by plotting in MAPLE!

2. $F(x_1, x_2) = x_1^2$ has global minimums at each point $x^* = (0, x_2)$, check by plotting in MAPLE!

1.2.1 First Order Conditions

Theorem 1 *If $x^* \in U \subset R^n$ is a local max or min, then x^* is a critical point, that is*

$$\frac{\partial F}{\partial x_i}(x^*) = 0, \quad i = 1, ..., n.$$

Proof. Let us assign to our n variable function the one variable function

$$f(x_i) = F(x_1^*, ..., x_{i-1}^*, x_i, x_{i+1}^*, ..., x_n^*).$$

The derivative $f'(x_i^*) = \frac{dF(x_1^*,...,x_{i-1}^*,x_i,x_{i+1}^*,...,x_n^*)}{dx_i}(x_i^*)$ coincides with the partial derivative $\frac{\partial F}{\partial x_i}(x_1^*, ..., x_n^*)$. The function $f(x_i)$ has a local optimum at x_i^*, thus $f'(x_i^*) = 0$, this completes the proof.
Example. Find critical points for $F(x, y) = x^3 - y^3 + 9xy$.
Solution.

$$\frac{\partial F}{\partial x}(x, y) = 3x^2 + 9y, \quad \frac{\partial F}{\partial y}(x, y) = -3y^2 + 9x.$$

To find critical points solve the system

$$\begin{cases} \frac{\partial F}{\partial x}(x, y) = 0 \\ \frac{\partial F}{\partial y}(x, y) = 0 \end{cases} \left| \begin{array}{c} 3x^2 + 9y = 0 \\ -3y^2 + 9x = 0 \end{array} \right| -\frac{1}{3}x^4 + 9x = 0$$

the solutions are $(x = 0, y = 0)$, $(x = 3, y = -3)$. To determine whether either of these critical points is min max or neither we need *second order conditions* which involve second derivatives of F.

1.2.2 Second Order Sufficient Conditions

We use the notation $\frac{\partial^2 F}{\partial x_i \partial x_j} = F_{x_i x_j}$. Let

$$D^2 F = \begin{pmatrix} F_{x_1x_1} & F_{x_2x_1} & ... & F_{x_nx_1} \\ F_{x_1x_2} & F_{x_2x_2} & ... & F_{x_nx_2} \\ ... & ... & ... & ... \\ F_{x_1x_n} & F_{x_2x_n} & ... & F_{x_nx_n} \end{pmatrix}$$

be the Hessian matrix of F.

Theorem 2 *Suppose $x^* \in U \subset R^n$ is a critical point, then*
(1) If $D^2F(x^*)$ **negative definite** *then x^* is* **strict local max***;*
(2) If $D^2F(x^*)$ **positive definite** *then x^* is* **strict local min***;*
(3) If $D^2F(x^*)$ **indefinite** *then x^* is a* **saddle point** *(that is a critical point which is neither a local max nor a local min).*

Sketch of Proof. Consider the approximation of F by it's second order Taylor polynomial

$$F(x^* + h) \approx F(x^*) + D^1F(x^*) \cdot h + \frac{1}{2} h^T \cdot D^2F(x^*) \cdot h.$$

Since x^* is a critical point $D^1F(x^*) = 0$, thus

$$F(x^* + h) - F(x^*) \approx h^T \cdot D^2F(x^*) \cdot h.$$

(1) If our quadratic form is negative definite, then for all $h \neq 0$ the right hand side is negative, thus so is the left hand side in some neighborhood of x^*, this completes the proof.

Similarly for (2) and (3).

Do you remember when a quadratic form is negative definite?
Let us recall.

Let $Q(x_1, ..., x_n) = (x_1, ..., x_n) \cdot \begin{pmatrix} a_{11} & ... & a_{1n} \\ a_{21} & ... & a_{2n} \\ ... & ... & ... \\ a_{n1} & ... & a_{nn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ ... \\ x_n \end{pmatrix}$ be an n variable quadratic form.

The following n determinants

$$|D_1| = \begin{vmatrix} a_{11} \end{vmatrix}, \quad |D_2| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \ ... \ , \ |D_n| = \begin{vmatrix} a_{11} & ... & a_{1n} \\ a_{21} & ... & a_{2n} \\ ... & ... & ... \\ a_{n1} & ... & a_{nn} \end{vmatrix}$$

are called *leading principal minors*.

Theorem 3 *A quadratic form is positive definite if and only if*

$$|D_1| > 0, \quad |D_2| > 0, \quad ... \quad , \quad |D_n| > 0$$

that is all principal minors are positive.

A quadratic form is negative definite if and only if

$$|D_1| < 0, \quad |D_2| > 0, \quad |D_3| < 0, \quad ...$$

that is principal minors alternate in sign starting with negative one.

From this definiteness criteria and the Theorem 2 follows

Corollary 1 *Suppose*

$$\frac{\partial F}{\partial x_i} = 0, \quad i = 1, 2, ..., n,$$

and n leading principal minors of $D^2F(x^*)$ alternate in sign

$$\left| F_{x_1x_1} \right| < 0, \quad \begin{vmatrix} F_{x_1x_1} & F_{x_2x_1} \\ F_{x_1x_2} & F_{x_2x_2} \end{vmatrix} > 0, \quad \begin{vmatrix} F_{x_1x_1} & F_{x_2x_1} & F_{x_3x_1} \\ F_{x_1x_2} & F_{x_2x_2} & F_{x_3x_2} \\ F_{x_1x_3} & F_{x_2x_3} & F_{x_3x_3} \end{vmatrix} < 0, \ ...$$

at x^*. Then x^* is a strict local max.

Corollary 2 *Suppose*

$$\frac{\partial F}{\partial x_i} = 0, \quad i = 1, 2, ..., n,$$

and n leading principal minors of $D^2F(x^*)$ are positive

$$\left| F_{x_1x_1} \right| > 0, \quad \begin{vmatrix} F_{x_1x_1} & F_{x_2x_1} \\ F_{x_1x_2} & F_{x_2x_2} \end{vmatrix} > 0, \quad \begin{vmatrix} F_{x_1x_1} & F_{x_2x_1} & F_{x_3x_1} \\ F_{x_1x_2} & F_{x_2x_2} & F_{x_3x_2} \\ F_{x_1x_3} & F_{x_2x_3} & F_{x_3x_3} \end{vmatrix} > 0, \ ...$$

at x^*. Then x^* is a strict local min.

Corollary 3 *Suppose*

$$\frac{\partial F}{\partial x_i} = 0, \quad i = 1, 2, ..., n,$$

and some nonzero leading principal minors of $D^2F(x^*)$ violate the sign patterns of Corollaries 1 and 2. Then x^* is a saddle point.

Example. Now we can classify two critical points $(0, 0)$ and $(3, -3)$ of the function $F(x, y) = x^3 - y^3 + 9xy$. The Hessian of F is

$$\begin{vmatrix} F_{xx} & F_{yx} \\ F_{xy} & F_{yy} \end{vmatrix} = \begin{vmatrix} 6x & 9 \\ 9 & -6y \end{vmatrix}.$$

The first leading principal minor is $F_{xx} = 6x$ and the second order principal leading minor is $|D^2F(x)| = -36xy - 81$.

At $(0, 0)$ these two minors are 0 and -81, this is the situation of Corollary 3, so $(0, 0)$ is a saddle point.

At $(3, -3)$ these two minors are 18 and 24, this is the situation of Corollary 2, so $(3, -3)$ is a strict local min point.

Note that this local min is not global: the restriction $F(0, y) = 0^3 - y^3 + 9 \cdot 0 \cdot y = -y^3$ decreases to $-\infty$ when y increases to ∞.

1.2.3 Necessary Conditions

For a one variable function we have the following implications:

$$x^* \text{ } is \text{ } min \quad \Rightarrow f'(x^*) = 0$$

that is "$f'(x^*) = 0$" is *necessary* condition for minimality;

$$x^* \text{ } is \text{ } min \quad \Leftarrow \begin{cases} f'(x^*) = 0 \\ f''(x^*) > 0 \end{cases}$$

that is "$f'(x^*) = 0$ *and* $f''(x^*) > 0$" is *sufficient* for minimality.

But not

$$x^* \text{ } is \text{ } min \quad \Rightarrow \begin{cases} f'(x^*) = 0 \\ f''(x^*) > 0 \end{cases}$$

that is "$f'(x^*) = 0$ *and* $f''(x^*) > 0$" is *not necessary*: counterexample is $f(x) = x^4$.

But

$$x^* \text{ } is \text{ } min \quad \Rightarrow \begin{cases} f'(x^*) = 0 \\ f''(x^*) \geq 0 \end{cases}$$

that is "$f'(x^*) = 0$ *and* $f''(x^*) \geq 0$" is *necessary.*

Is the last condition *sufficient* for minimality? In other words, is the implication

$$x^* \text{ } is \text{ } min \quad \Leftarrow \begin{cases} f'(x^*) = 0 \\ f''(x^*) \geq 0 \end{cases}$$

correct?

No: let $f(x) = x^4$, $g(x) = -x^4$, $h(x) = x^3$, the point $x^* = 0$ is critical for all these functions, besides

$$f''(0) = g''(0) = h''(0) = 0 \geq 0,$$

but $x^* = 0$ is a **min** for f, **max** for g and **neither** for h.

A similar result is true in multivariable case

Theorem 4 *Suppose x^* is a local min (max) for F. Then, $\nabla F(x^*) = 0$ and $D^2 F(x^*)$ is negative (positive) semidefinite.*

Corollary 4 *(a) If x^* is a local min for F, then*

$$\frac{\partial F}{\partial x_i}(x^*) = 0, \quad i = 1, ..., n,$$

and all the principal minors of Hessian $D^2 F(x^*)$ are ≥ 0.

(b) If x^* is a local max for F, then

$$\frac{\partial F}{\partial x_i}(x^*) = 0, \quad i = 1, ..., n,$$

and all the principal minors of Hessian $D^2 F(x^*)$ of odd degree are ≤ 0 and all principal minors of even degree are ≥ 0.

1.3 Global Maxima and Minima

How to recognize that a local max (min) found by the above described first and second condition is global?

In one variable case there are such criteria:

1. If x^* is a local max (min) and it is the *only* critical point, then it is global.

2. If $f''(x) \leq 0$ ($f''(x) \geq 0$) for *each* x and x^* is a local max (min), then it is global.

The first condition does work in higher dimensions, but the condition 2 *does*:

Theorem 5 *If x^* is a critical point $\nabla F(x^*) = 0$, and the Hessian $D^2F(x)$ is negative (positive) semidefinite* **for all** *x, then x^* is global max (min).*

1.4 Economic Applications

1.4.1 Profit-Maximizing Firm

Suppose a firm uses n inputs and produces one output, and the production function is given by $G(x)$, here x is input bundle $x = (x_1, ..., x_n)$. Suppose the price of the unit of output product is p. The revenue then is $R(x) = pG(x)$, and the profit is

$$F(x) = R(x) - C(x) = pG(x) - C(x),$$

where $C(x)$ is the cost function.

For which value of input bundle x the profit is maximal?

The first order necessary condition gives

$$0 = \frac{\partial F}{\partial x_i}(x) = \frac{\partial R}{\partial x_i}(x) - \frac{\partial C}{\partial x_i}(x),$$

that is where marginal revenue equals to marginal cost.
Question: Can you explain the economical meaning of this fact?

Consider now more simple situation - **constant marginal cost**: suppose $C(x) = w \cdot x = w_1x_1 + ... + w_nx_n$. In this case

$$\frac{\partial C}{\partial x_i}(x) = w_i.$$

On the other hand

$$\frac{\partial R}{\partial x_i}(x) = \frac{\partial pG}{\partial x_i}(x) = p\frac{\partial G}{\partial x_i}(x),$$

so the first order condition $\frac{\partial R}{\partial x_i}(x) = \frac{\partial C}{\partial x_i}(x)$ gives

$$p\frac{\partial G}{\partial x_i}(x) = w_i, \quad thus \quad \frac{\partial G}{\partial x_i}(x) = \frac{w_i}{p}.$$

So to be a maximizer pretends a solution x^* of the equation

$$\frac{\partial G}{\partial x_i}(x) = \frac{w_i}{p}.$$

Now look at the second order condition. By necessary condition, for maximizer x^* the Hessian $D^2F(x^*)$ must be negative semidefinite. In our case of constant marginal cost, easy to see that

$$D^2(F)(x) = pD^2G(x).$$

Thus for maximizer the Hessian $D^2G(x^*)$ must be negative semidefinite. In particular all principal minors of degree 1 must be nonpositive, that is $\frac{\partial^2 G}{\partial x_i^2} \leq 0$. This is necessary condition for a critical point x^* to be maximizer.

1.4.2 Discriminating Monopolist

Suppose a firm sells its product at two markets (say domestic and foreign). Let Q_1 be the amount supplied to market 1 and Q_2 be the amount supplied to market 2. Suppose the inverse price functions are

$$P_1 = G_1(Q_1), \quad P_2 = G_2(Q_2).$$

What is revenue? $R(Q_1 + Q_2) = R(Q_1) + R(Q_2)$, where

$$R(Q_1) = Q_1P_1 = Q_1G(Q_1), \quad R(Q_2) = Q_2P_2 = Q_2G(Q_2).$$

Suppose that the total production cost C depends on the total production $Q_1 + Q_2$, so $C = C(Q_1 + Q_2)$. Then the profit is

$$\begin{gathered} F(Q_1, Q_2) = R(Q_1, Q_2) - C(Q_1, Q_2) = \\ R(Q_1) + R(Q_2) - C(Q_1, Q_2) = Q_1G(Q_1) + Q_2G(Q_2) - C(Q_1, Q_2). \end{gathered}$$

The first order condition gives that a maximizer must satisfy

$$\begin{array}{c} 0 = \frac{\partial F}{\partial Q_1} = \frac{dQ_1G(Q_1)}{dQ_1} - C'(Q_1 + Q_2); \\ 0 = \frac{\partial F}{\partial Q_2} = \frac{dQ_2G(Q_2)}{dQ_2} - C'(Q_1 + Q_2) \end{array}$$

or

$$\frac{dQ_1G(Q_1)}{dQ_1} = \frac{dQ_2G(Q_2)}{dQ_2} = C'(Q_1 + Q_2),$$

So the marginal revenue in each market should equal the marginal cost of total output.

Example. A monopolist producing a single output has two types of customers. If it produces Q_1 units for customers of type 1, then these customers are willing to pay a price of $50 - 5Q_1$ dollars per unit. If it produces Q_2 units for customers of type 2, then these customers are willing to pay a price of

$100-10Q_2$ dollars per unit. The monopolist's cost of manufacturing Q units of output is $90+20Q$ dollars. In order to maximize profits, how much should the monopolist produce for each market?
Solution. The profit function is

$$F(Q_1,Q_2)=$$

$$Q_1(50-5Q_1)+Q_2(100-lOQ_2)-(90+20(Q_1+Q_2)).$$

The critical points of F satisfy

$$\frac{\partial F}{\partial Q_1}=50-10Q_1-20=0,\quad Q_1=3,$$
$$\frac{\partial F}{\partial Q_2}=100-20Q_2-20=0,\quad Q_2=4.$$

So the critical point is $(3,4)$.

Now check the second order conditions. Since

$$F_{Q_1Q_1}=-10,\quad F_{Q_2Q_2}=-20,\quad F_{Q_1Q_2}=0,$$

the Hessian looks as

$$D^2(Q_1,Q_2)=\begin{pmatrix} F_{Q_1Q_1} & F_{Q_2Q_1} \\ F_{Q_1Q_2} & F_{Q_2Q_2} \end{pmatrix}=\begin{pmatrix} -10 & 0 \\ 0 & -20 \end{pmatrix}.$$

The first order leading principal minor of $D^2F(3,4)$ is -10 and the second leading principal minor is 200. Therefore $(3,4)$ is strict local max. But is it a global max?

Yes, it is, because the Hessian $D^2F(x)$ is a constant (that is independent on x) negative definite matrix.

1.5 Application: Regression Analysis - Least Squares Approximation

Regression analysis is the process of fitting an elementary function to a set of data points using the method of least squares.

1.5.1 Linear Case

Suppose the points $(x_i,y_i)\in R^2,\ i=1,2,...,n$ are given. Aim is to construct a linear function $y=ax+b$ that best fits these data points.

Idea: for each x_i consider the difference y_i-ax_i-b (called residual), and take the sum of squares of all residuals

$$F(a,b)=\sum_{i=1}^{n}(y_i-ax_i-b)^2,$$

this is a function of two variables, a and b. The goal is to minimize this function. **For this just solve the system of equations $F_a=0,\ F_b=0$ for a and b.**

Question: Why the sum o **squares**, why not just the sum of all residuals, or the sum of absolute values of all residuals?

Formal (boring) Solution.

$$F_a = \sum_{i=1}^{n} 2(y_i - ax_i - b)(-x_i) = 0,$$
$$F_b = \sum_{i=1}^{n} 2(y_i - ax_i - b)(-1) = 0.$$

Simplifying we obtain the system

$$\begin{cases} (\sum_{i=1}^{n} x_i^2)a & + & (\sum_{i=1}^{n} x_i)b & = & \sum_{i=1}^{n} x_i y_i \\ (\sum_{i=1}^{n} x_i)a & + & nb & = & \sum_{i=1}^{n} y_i \end{cases}.$$

Solving this system we obtain

$$a = \frac{n\cdot(\sum_{i=1}^{n} x_i y_i) - (\sum_{i=1}^{n} x_i)\cdot(\sum_{i=1}^{n} y_i)}{n\cdot(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2}$$
$$b = \frac{(\sum_{i=1}^{n} x_i^2)\cdot(\sum_{i=1}^{n} y_i) - (\sum_{i=1}^{n} x_i)\cdot(\sum_{i=1}^{n} x_i y_i)}{n\cdot(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2}$$

Example. Fit the data $(2, 4), (5, 6), (6, 7), (9, 8)$ by linear function.
Solution. Let us try to approximate by

$$y = ax + b.$$

The function F then looks as

$$F(a, b) = (4 - (2a + b))^2 + (6 - (5a + b))^2 + (7 - (6a + b)^2 + (8 - (9a + b)^2.$$

Simplification gives

$$F(a, b) = 165 - 304a - 50b + 146a^2 + 44a * b + 4b^2.$$

The partials of F are

$$F_a = -304 + 292a + 44b, \quad F_b = -50 + 44a + 8b.$$

Solution of the system

$$\begin{cases} -304 + 292a + 44b = 0 \\ -50 + 44a + 8b = 0 \end{cases}$$

gives $a = 0.59, \quad b = 3.06$, so the best fitting line is

$$y = 0.59x + 3.06.$$

Now we solve the same problem using above formulae:

x_i	y_i	$x_i y_i$	x_i^2
2	4	8	4
5	6	30	25
6	7	42	36
9	8	72	81
22	25	152	146,

so

$$a = \frac{4 \cdot 152 - 22 \cdot 25}{4 \cdot 146 - 22^3} = 0.58,$$
$$b = \frac{146 \cdot 25 - 22 \cdot 152}{4 \cdot 146 - 22^3} = 3.06,$$

thus the best fitting line is

$$y = 0.58x + 3.06.$$

MAPLE makes it faster:

$> with(Statistics):$
$> X := Vector([2, 5, 6, 9]):$
$> Y := Vector([4, 6, 7, 8]):$
$> Fit(a * x + b, X, Y, x);$
$0.579999999999999960x + 3.05999999999999828$

1.5.2 Quadratic Case

Suppose the points $(x_i, y_i) \in R^2,\ i = 1, 2, ..., n$ are given. Aim is to construct a quadratic function $y = ax^2 + bx + c$ that best fits these data points.

Idea: for each x_i consider the difference $y_i - ax_i^2 - bx_i - c$ (called residual), and take the sum of squares of all residuals

$$F(a, b) = \sum_{i=1}^{n} (y_i - ax_i^2 - bx_i - c)^2,$$

this is a function of three variables, $a,\ b$ and c. The goal is to minimize this function. **For this just solve the system of equations $F_a = 0,\ F_b = 0,\ F_c = 0$ for $a,\ b$ and c.**

Same for higher order polynomials.

Exercises

1. For each of the following functions find the critical points and classify these as local max, local min, saddle point, or "can't tell" using first and second order conditions:
(a) $F(x,y) = x^2 + xy + y^2 - 3x$, (b) $F(x,y) = xy - x^3 - y^2$;
(c) $F(x,y) = xy^2 + x^3y - xy$, (d) $F(x,y) = 3x^4 + 3x^2y - y^3$.

2. A firm produces two kind of golf ball, one that sells for \$3 and one for \$2. The total cost, in thousands of dollars, of producing of x thousand balls at \$3 each and y thousand balls at \$2 each is given by

$$C(x,y) = 2x^2 - 2xy + y^2 - 9x + 6y + 7.$$

Find the amount of each type of ball that must be produced and sold in order to maximize profit.

3. A one-product company finds that its profit, in millions of dollars, is a function P given by

$$P(a,p) = 2ap + 80p - 15p^2 - 1/10 \cdot a^2p - 100,$$

where a is the amount spent on advertising, in millions of dollars, and p is the price charged per item of the product, in dollars. Find the maximum value of P and the values of a and p at which it is attained.

4. A one-product company finds that its profit, in millions of dollars, is a function P given by

$$P(a,n) = -5a^2 - 3n^2 + 48a - 4n + 2an + 300,$$

where a is the amount spent on advertising, in millions of dollars, and n is the number of items sold, in thousands. Find the maximum value of P and the values of a and n at which it is attained.

5. A trash company is designing an open-top, rectangular container that will have a volume of 320 ft^3. The cost of making the bottom of the container is \$5 per square foot, and the cost of the sides is \$4 per square foot. Find the dimensions of the container that will minimize total cost. (Hint: Make a substitution using the formula for volume.)

6. A computer firm, markets two kinds of electronic calculator that compete with one another. Their demand functions are expressed by the following relationships:

$$q_1 = 78 - 6p_1 - 3p_2,$$

$$q_2 = 66 - 3p_1 - 6p_2,$$

where p_1 and p_2 are the price of each calculator, in multiples of \$10, and q_l and q_2 are the quantity of each calculator demanded, in hundreds of units.

a) Find a formula for the total-revenue function R in terms of the variables p_1 and p_2.

b) What prices p_1 and p_2 should be charged for each product in order to maximize total revenue?

c) How many units will be demanded?

d) What is the maximum total revenue?

7. For the following data

$$(0, 10), (5, 22), 10, 31), (15, 46), (20, 51)$$

find the last squares line and estimate y when $x = 2.5$.

8. Find the coefficients of the parabola

$$y = ax^2 + bx + c$$

that is the best fit for the points

$$((1, 2), (2, 1), (3, 1), (4, 3).$$

Exercise 17.1-17.8 from [SB].