- For another source with $Q = \{0.6180, 0.3820\}$, its transmission rate is

$$R(Q) = \frac{H(Q)}{T(Q)} = \frac{0.618\lg(1/0.618) + 0.382\lg(1/0.382)}{0.618(1) + 0.382(2)} = 0.6942 \text{ bps}$$

It shows that a source can go through the channel at a faster or slower rate than the mean rate. The rate 0.6942 bps is about the fastest this channel is capable of, the subject of next two sections.

Exercises 1.4

1. Let $Q = \{1/4, 1/4, 1/4, 1/4\}$ be the distribution of an information source, $C = \{1, 10, 100, 1000\}$ be an encoding scheme to the binary system. Find the mean rates $R(Q), R(P)$ when Q and $B = \{0, 1\}$ are treated as different channels.
2. Consider the Morse code of Example 1.4.1. Find the mean rate if the timings of the symbols are changed to: $\tau_{-} = 2\tau_{\bullet}$, $\tau_{\wedge} = \tau_{\bullet}$, $\tau_{/} = 2\tau_{\bullet}$, $\tau_{\sqcup} = 3\tau_{\bullet}$).
3. Prove Theorem 1.16.
4. Consider a comma code of 4 code words, $1, 10, 100, 1000$. Assume each binary symbol takes the same amount of time to transmit, t_0. Find the transmission rate $R(P)$ of a source whose probability distribution is $P = \{2/5, 3/10, 1/5, 1/10\}$.

1.5 LAGRANGE MULTIPLIER METHOD

A Prototypic Example — Get Out the Hot Spot

Consider an experiment consisting of an oval metal plate, a heat source under the plate, and a bug on the plate. A thought experiment is sufficient so that no actual animals are harmed. In one setup, Fig.1.3(a), the heat source is place at the center of the plate, and the bug is placed at a point as shown. Assume the heat source is too intense for comfort for the bug. Where will it go? By inspection, you see the solution right away. The coolest spots are the edge points through plate's major axis, and the bug should go to whichever is closer. However, the bug cannot see the "whole" picture. It has its own way to solve the problem. Translating the bug's solution into mathematics gives rise to the Lagrange Multiplier Method. We break its solution into two steps: the solution to reach plate's edge, and the solution to go to the coolest point on the edge.

Level Curve and Gradient

Place the plate in the coordinate system as shown. Let (x_0, y_0) be the bug's current position as shown. Let $T(x, y)$ denote the temperature in $°F$

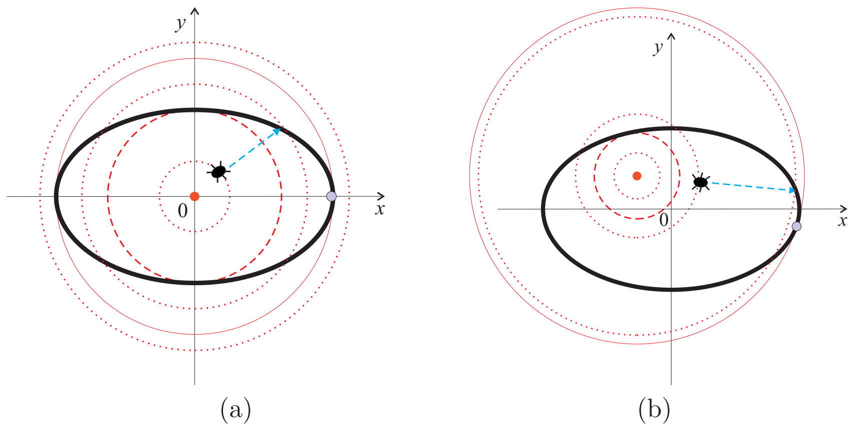

Figure 1.3 Thought experiment set-up.

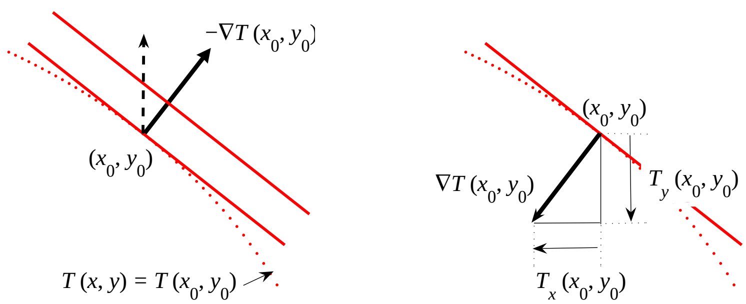

Figure 1.4 Dashed arrow is a non-gradient vector.

at any point (x, y) in the plate. So the bug experiences a temperature of $T(x_0, y_0)$ °F. The bug's immediate concern is to move away from the heat source as fast as possible. The direction it takes is denoted by $-\nabla T(x_0, y_0)$ and the opposite direction $\nabla T(x_0, y_0)$ is called the gradient of T at the point.

To derive the gradient, let us give function T a specific form for illustration purposes. For simplicity, we assume that T is distributed concentrically, i.e., the temperature remains constant in any concentric circle around the heat source $(0, 0)$, and the closer to the source, the higher the constant temperature. Such curves are called isothermal curves for this situation and **level curves** in general. Obviously, the bug will not go into its current isothermal curve $T(x, y) = T(x_0, y_0)$, nor stay on it, walking in circle so to speak. It must go to another level curve whose isothermal temperature is immediately lower, $T(x, y) = k$ with $k < T(x_0, y_0)$. If we zoom in on the graphs, we see the answer quickly. At close-up, any smooth curve looks like a line, see Fig.1.4, the **tangent line** by definition. In fact, all tangent lines to level curves nearby are packed like parallel lines. Therefore, the shortest path from the tangent line of the current level curve to the tangent line of another level curve nearby is through the line perpendicular to these parallel tangent lines. The perpendicular direction leading to higher values of T is the **gradient** $\nabla T(x_0, y_0)$, and the opposite direction leading to lower values of T is $-\nabla T(x_0, y_0)$.

Next, we derive the direction analytically. Let $y = y(x)$ represent the level curve $T(x, y) = T(x_0, y_0)$ near (x_0, y_0), i.e., $T(x, y(x)) \equiv T(x_0, y_0)$ for all x near x_0 and $y(x_0) = y_0$. Then $\frac{dy}{dx}(x_0)$ is the slope of the tangent line $y = y_0 + \frac{dy}{dx}(x_0)(x - x_0)$. The slope is found by the Chain Rule and Implicity Differentiation. Formally we have,

$$\begin{aligned}
\frac{d}{dx}T(x,y) = \frac{d}{dx}T(x_0,y_0) &= 0 \\
T_x(x,y) + T_y(x,y)\frac{dy}{dx} &= 0 \\
\frac{dy}{dx} &= -T_x(x,y)/T_y(x,y) \\
\frac{dy}{dx}(x_0) &= -T_x(x_0,y_0)/T_y(x_0,y_0),
\end{aligned}$$

where $T_x(x, y) = \partial T(x, y)/\partial x$ is the partial derivative of T with respect to variable x, and similarly for $T_y(x, y)$. Since the gradient direction is perpendicular to the tangent of the level curve, its slope, m, is the negative reciprocal of the tangent slope,

$$m = -\frac{1}{dy/dx} = \frac{T_y(x_0,y_0)}{T_x(x_0,y_0)},$$

for which $T_x(x_0, y_0)$ can be thought to be a run of the gradient line and $T_y(x_0, y_0)$ its rise. Finally, we have

---

Definition 1.17 *Let $F(x, y)$ be a differentiable function, the **gradient** of F at (x, y) is the vector*

$$\nabla F(x,y) = (F_x(x,y), F_y(x,y)).$$

Important Properties of Gradient

1. Function value increases the fastest in the gradient direction.
2. Function value decreases the fastest in the opposite gradient direction.
3. The gradient is perpendicular to its level curve.

Constrained Optimal Solutions

Driven by instinct, the bug runs in the negative gradient direction to escape the heat. This strategy will only lead it to the edge of the plate, see Fig.1.3(a). Its next move is *constrained* by the boundary curve of the plate. In one direction, heat increases to its highest on the y-axis. In the other, heat decreases to the lowest on the x-axis. The critical clue for the bug as to which direction to move next when reaching the boundary is the fact that the isothermal curve is transversal to the boundary curve. At the minimum temperature point, both the isothermal curve and the boundary curve are tangent, and moving in either direction does not lower the heat further.

For the situation of Fig.1.3(a), we see the exact solution right away. We are even able to derive the solution *graphically* in general such as Fig.1.3(b). That is, when the first isothermal curve, radiating outwards from the source, touches the boundary it gives rise to the constrained maximum temperature point, and when the last curve to leave the boundary it gives rise to the constrained minimum temperature point. The question we address below is how to translate this picture into *analytical* terms.

To this end, let us derive the analytical method for the trivial case first. To be specific, let $\frac{x^2}{9}+\frac{y^2}{4}=1$ be the boundary curve of the plate so that $(3,0)$ is the constrained coolest point that the bug comes to rest. We further introduce a notation $g(x,y)=\frac{x^2}{9}+\frac{y^2}{4}$. Thus the boundary can be thought as the level curve of the function g: $g(x,y)=1$, and the plate the region: $g(x,y)\le 1$. Therefore the condition for both level curves, $T(x,y)=T(3,0)$ and $g(x,y)=1$, to be tangent at $(3,0)$ is to have the same gradient slope:

$$\frac{T_y(3,0)}{T_x(3,0)}=\frac{g_y(3,0)}{g_x(3,0)}.$$

An equivalent condition is

$$T_x(3,0)=\lambda g_x(3,0),\quad T_y(3,0)=\lambda g_y(3,0),$$

where λ is a scaler parameter. In fact, if $g_x(3,0)\ne 0$, $\lambda=T_x(3,0)/g_x(3,0)$ as one can readily check. Hence, the constrained optimal point $(3,0)$ is a solution to the equations

$$T_x(x,y)=\lambda g_x(x,y),\quad T_y(x,y)=\lambda g_y(x,y),\quad g(x,y)=1. \tag{1.4}$$

This problem has three unknowns x,y,λ and three equations. For the trivial case, it should have 4 solutions: 2 constrained maximums and 2 constrained minimums. The Lagrange Multiplier Method is to solve these equations for the extrema of function $T(x,y)$ subject to the constraint $g(x,y)=1$.

Example 1.5.1 The heat source is located to a new point $(0,1)$. Let $T(x,y)=\frac{1000}{1+x^2+(y-1)^2}$ be the temperature. (It is easy to check that the isothermal curves are concentric circles around the source $(0,1)$.) The constrained extrema (x,y) satisfy equations (1.4), which in this particular case are

$$T_x=-\frac{2000x}{[1+x^2+(y-1)^2]^2}=\lambda g_x(x,y)=\lambda\frac{2}{9}x$$

$$T_y = -\frac{2000(y-1)}{[1+x^2+(y-1)^2]^2} = \lambda g_y(x,y) = \lambda \frac{2}{4} y$$
$$g(x,y) = \frac{x^2}{9} + \frac{y^2}{4} = 1$$

Divide the first equation by the second to eliminate λ,

$$\frac{x}{y-1} = \frac{4}{9}\frac{x}{y}.$$

This equation has 2 types of solutions: (1) $x = 0$, then $y = \pm 2$. (2) $x \neq 0$, then $1/(y-1) = 4/(9y)$. Solving it gives $y = -4/5$ and $x = \pm 3\sqrt{1-y^2/4} = \pm 3\sqrt{21}/5$. Last, run a comparison contest among the candidate points, we find that $(0,2)$ is the constrained maximum with $T = 500°F$, $(\pm 3\sqrt{21}/5, -4/5)$ are the constrained minimums with $T = 80.5°F$. Point $(0,-2)$ is a constrained local maximum with $T = 100°F$.

Generalization — Lagrange Multiplier Method

In general, let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ denote an n-vector variable, $f(\mathbf{x})$ be a real-valued function. Then the n-vector

$$\nabla f(\mathbf{x}) = (f_{x_1}(\mathbf{x}), f_{x_2}(\mathbf{x}), \ldots, f_{x_n}(\mathbf{x}))$$

is the **gradient** of f at $\mathbf{x}$. In the gradient direction f increases most rapidly and opposite the gradient f decreases most rapidly. In addition, the gradient vector $\nabla f(\mathbf{x}_0)$ is perpendicular to the hypersurface (level surface) $f(\mathbf{x}) = f(\mathbf{x}_0)$.

A constrained optimization problem is to find either the maximum value or the minimum value of a function $f(\mathbf{x})$ with $\mathbf{x}$ constrained to a level hypersurface $g(\mathbf{x}) = k$ for some function g and constant k. The following theorem forms the basis of the Lagrange Multiplier Method.

Theorem 1.18 *Suppose that both f and g have continuous first partial derivatives. If either*

- *the maximum value of $f(\boldsymbol{x})$ subject to the constraint $g(\boldsymbol{x}) = 0$ occurs at $\boldsymbol{x}_0$; or*
- *the minimum value of $f(\boldsymbol{x})$ subject to the constraint $g(\boldsymbol{x}) = 0$ occurs at $\boldsymbol{x}_0$,*

then $\nabla f(\boldsymbol{x}_0) = \lambda \nabla g(\boldsymbol{x}_0)$ for some constant λ.

To find the candidate points $\mathbf{x}_0$ for the constrained extrema, the method calls to solve the system of equations

$$\begin{cases} \nabla f(\mathbf{x}) = \lambda \nabla g(\mathbf{x}) \\ g(\mathbf{x}) = 0. \end{cases} \tag{1.5}$$

There are $n+1$ variables $(\mathbf{x}, \lambda)$ and $n+1$ equations since the vector equation for the gradients has n scalar equations. We expect to have non-unique solutions since they are candidates for both constrained maximum and constrained minimum.

Example 1.5.2 We now give an alternative proof to Theorem 1.11. It is to solve the constrained optimization problem

$$\begin{aligned} \text{Maximize: } & H(p) = -(p_1 \lg p_1 + p_2 \lg p_2 + \cdots + p_n \lg p_n) \\ \text{Subject to: } & g(p) = p_1 + p_2 + \cdots p_n = 1. \end{aligned}$$

By Lagrange Multiplier Method,

$$\begin{aligned} \nabla H(p) &= \lambda \nabla g(p) \\ g(p) &= 1 \end{aligned}$$

In component, $H_{p_k} = -(\lg p_k + 1) = \lambda g_{p_k} = \lambda$. Hence, $p_k = 2^{-\lambda-1}$ for all k. Since $2^{-\lambda-1}$ is a constant for any λ, the distribution must be the equiprobability distribution $p_k = 1/n$. Since there are distributions, for example $p = (1, 0, \ldots, 0)$, at which $H = 0$, and since the equiprobability distribution is the only constrained solution, it must be the constraint maximum because its value is positive $H = \lg n > 0$.

Exercises 1.5 (Give numerical approximations if exact solutions are not feasible.)

1. Consider the prototypic example from the main text. Suppose the heat source is relocated to a new point $(1, 1)$ and $T(x, y) = \frac{1000}{1+(x-1)^2+(y-1)^2}$ is the temperature function. Assume the same boundary $g(x, y) = \frac{x^2}{9} + \frac{y^2}{4} = 1$ for the metal plate. Find both the hottest and coolest spots on the plate.
2. Find the minimum distance between the point $(3, 3)$ and the ellipse $x^2/4 + y^2/9 = 1$.
3. The Lagrange Multiplier Method can be generalized to multiple constraints. For example, to find optimal values of a function $f(\mathbf{x})$ subject to two constraints $g(\mathbf{x}) = 0, h(\mathbf{x}) = 0$, we solve the following equations
$$\begin{cases} \nabla f(\mathbf{x}) = \lambda \nabla g(\mathbf{x}) + \mu \nabla h(\mathbf{x}) \\ g(\mathbf{x}) = 0 \\ h(\mathbf{x}) = 0 \end{cases}$$
for candidate extrema. Here λ and μ are multipliers. Let an ellipse be the intersection of the plane $x + y + z = 4$ and the paraboloid $z = x^2 + y^2$. Find the point on the ellipse that is closest to the origin.
4. **The Least Square Method:** Let $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be a collection of data points. Assume x is the independent variable and y is the dependent variable, and we wish to fit the data to a function $y = f(x)$ from a class of functions. Each possible fitting function has a deviation from each data point, $d_k = y_k - f(x_k)$. Function f is a **least square fitting** function if
$$\sum_{k=1}^{n} d_k^2 = \sum_{k=1}^{n} [y_k - f(x_k)]^2 = \text{ a minimum}$$

in the class. Here $\sum d_k^2$ can be thought as the square of a distance between the data points and the candidate function, and hence the name for the method. For example, if we wish to fit the data to a function of the linear class $y = f(x) = b + mx$ with b the y-intercept, m the slope to be determined, we want to minimize the following function of b, m:

$$F(b,m) = \sum_{k=1}^{n} [y_k - f(x_k)]^2 = \sum_{k=1}^{n} [y_k - (b + mx_k)]^2.$$

Use the Lagrange Multiplier Method to show that the least square fit satisfies

$$b = \frac{\sum x_k^2 \sum y_k - \sum x_k \sum x_k y_k}{n \sum x_k^2 - (\sum x_k)^2}, \quad m = \frac{n \sum x_k y_k - \sum x_k \sum y_k}{n \sum x_k^2 - (\sum x_k)^2}.$$

(*Hint:* Use the trivial constraint $g(b,m) \equiv 0$.)

5. Let $(1, 1.5), (2, 2.8), (3, 4.5)$ be three data points. Use the formula above to find the least square fit line $y = b + mx$.

1.6 OPTIMAL SOURCE RATE — CHANNEL CAPACITY

For a channel S, the mean rate R_n is a measure for all sources on average. But a particular source P can go through the channel at a slower or faster rate $R(P)$ than the mean rate. Once an optimal channel is chosen with respect to R_n, individual source can take advantage of the channel to transmit at a rate as fast as possible.

Definition 1.19 *The fastest source transmission rate over all possible sources is called the **channel capacity** of the channel. Denote it by $K = \max_P R(P)$.*

The channel capacity indeed exists and it is unique.

Theorem 1.20 *The source transmission rate $R(P)$ has a unique maximum K constrained to $\sum p_k = 1$. For the optimal source distribution, p_k^{1/τ_k} is a constant for all k, and $K = -\lg p_1/\tau_1 = -\lg p_k/\tau_k$. In particular, $p_k = p_1^{\tau_k/\tau_1}$, $\sum_{k=1}^n p_1^{\tau_k/\tau_1} = 1$.*

Proof. We use the Lagrange Multiplier Method to maximize $R(P) = H(P)/T(P)$ subject to the constraint $g(P) = \sum_{k=1}^n p_k = 1$. This is to solve the joint equations:

$$\begin{cases} \nabla R(P) = \lambda \nabla g(P) \\ g(P) = 1. \end{cases}$$

The first n component equations are

$$R_{p_k} = \frac{H_{p_k} T - H T_{p_k}}{T^2} = \lambda g_{p_k} = \lambda, \ k = 1, 2, \ldots, n.$$