4.1 Generalization and Specialization

Suppose that we want to solve a problem. One common first step is to study the problem in special cases, so that we understand the problem better—this process is called specialization. Another approach is to generalize the problem, studying a more general problem. You might think that generalizing makes the problem harder, but in several cases, a correct generalization will help us get rid of irrelevant information, and we can clearly see the entire picture.

An example. Suppose that our problem involves a property of triangles. To specialize this problem, we can, for example, consider the case where the triangle is equilateral. There are also several possible directions to generalize the problem. For example, we can generalize this property for polygons, or for tetrahedrons. See Figure 4.1 for an illustration. Note that generalization is mere guessing, and it is well possible that the generalized statement does not hold.

Pythagorean theorem. To illustrate the use of the methods above, consider the classical Pythagorean theorem. Geometrically, it states that on any right triangle, the total area of the two squares on the legs is the same as the area of the square on the hypotenuse. In other words, if the hypotenuse has length $a$, and the two legs have length $b$ and $c$ respectively, then $a^2 = b^2 + c^2$.

Now how should we generalize this problem? The theorem is involved with squares, so can we generalize it to polygons then? Of course the polygons must have some common property, otherwise you can find trivial counter-examples. In Pythagorean theorem, the polygons are of the same shape, specifically, they are all
Figure 4.2: Illustration of the Pythagorean theorem and its generalization.

squares. So let’s generalize the theorem such that instead of 3 squares, we have 3 similar polygons, meaning that those polygons are of the same shape. See Figure 4.2 for an illustration.

The generalized result looks complicated, but in fact it is equivalent to the classical Pythagorean theorem. To see why, suppose that the area of the polygon on the hypotenuse is $\lambda a^2$. Because the polygons are similar, the area of the polygons on the legs is $\lambda b^2$ and $\lambda c^2$ respectively. The generalized result claims that $\lambda a^2 = \lambda b^2 + \lambda c^2$, for any real number $\lambda > 0$. Even better, if we can prove the generalized result for a particular shape (instead of all possible shapes), then we also obtain the classical Pythagorean theorem.

Given that we have the freedom to choose the shape, it’s now time for specialization. The simplest polygon shape is certainly a triangle, but what triangle? Looking back in the picture of the classical Pythagorean theorem, there’s already a triangle in front of our noses, so maybe it’s our shape then?

In more details, let $ABC$ be the original right triangle, with $BC$ as the hypotenuse. Now, on $BC$, we draw a triangle of the same shape as $ABC$; this coincides with $ABC$ itself. The triangles on the legs are $ABH$ and $ACH$, as shown in Figure 4.3, where $AH$ is the altitude of $ABC$. The specialized theorem claims that the total area of $ABH$ and $ACH$ is the area of $ABC$, which is obviously true.

4.2 Another Example: Lines In The Plane

Suppose that we have $n$ lines in a plane in a “general position”, meaning that no two lines are parallel, and no three lines meet at the same point. We’d like to know how many regions those lines form. Let’s call this
number $F(n)$.

Let’s try to generalize this problem. What we have here is a problem in 2-dimensional space, so maybe we can generalize it to 3-dimensional space as follows: Suppose that we have $n$ planes in a general position, how many parts of the space do those planes form? We can even generalize the problem to $k$-dimensional space. At this point, the generalized problem seems to be quite scary, so let’s specialize this general problem with $k = 1$: For $n$ points in a line, how many segments do we have? See Figure 4.4 for an illustration. Let’s call this number $G(n)$. One can easily find $G(n) = n + 1$.

![Figure 4.4](image.png)

Figure 4.4: The specialized problem for 1-dimensional space, illustrated for the case $n = 3$.

Now let’s get back to the problem of computing $F(n)$. Let’s compute $F(n)$ for a few first values. The numbers are given in the table below.

<table>
<thead>
<tr>
<th>$n$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G(n)$</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>$F(n)$</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>11</td>
<td>16</td>
</tr>
</tbody>
</table>

The function $F(n)$ grows much faster than $G(n)$, but its rate of growth $F(n + 1) - F(n)$ seems to be exactly $G(n)$. So we suspect that we have found a rule $F(n + 1) = F(n) + G(n) = F(n) + n + 1$ for every $n$. (There are many ways to solve this recurrence. This is left as an exercise.)

![Figure 4.5](image.png)

Figure 4.5: Illustration for the recurrence $F(n + 1) = F(n) + G(n)$, for the case $n = 3$.

Let’s try to understand why the recurrence $F(n + 1) = F(n) + G(n)$ holds. Suppose that we have $n$ lines in a general position, and they form $F(n)$ regions. Now let’s draw a new line. This intersects the prior $n$ lines, forming $G(n) = n + 1$ segments. But these segments correspond to the new $n + 1$ regions that the new line creates, so it’s why $F(n + 1) = F(n) + G(n)$. See Figure 4.5 for an illustration.

Here we don’t solve the generalized problem in $k$-dimension, but you can also find their recurrences using the argument above.

**Exercise:** Two men are seated at a table of usual rectangular shape. One places a penny on the table, then the other does the same, and so on, alternately. It is understood that each penny lies flat on the table and not on any penny previously placed. The player who puts the last coin on the table takes the money. Which player should win, provided that each plays the best possible game?
4.3 Yet Another Example: Polygons

In many cases, it suffices to specialize. As an illustrative example, consider the following problem. Let $P$ be a convex polygon that is contained in another polygon $Q$, as illustrated in the left panel of Figure 4.6. It’s obvious that $P$’s area is smaller than that of $Q$, but what about their perimeter? We claim that the perimeter of $P$ is shorter than the perimeter of $Q$. This problem is however subtle. It’s only true if $P$ is convex. If $P$ is not convex then its perimeter can be as big as we want to, despite its confinement within $Q$. So the problem is not obvious at all, and our solution needs to somehow make use of the fact that $P$ is convex.

![Figure 4.6: Illustration of polygons $P$ and $Q$, in general position.](image)

Given that the convexity of $P$ plays an important role here, let’s first review the definition of convex polygons. Below are some common (equivalent) definitions:

- The polygon is entirely contained in a closed half-plane defined by each of its edges.
- Every internal angle is strictly less than 180 degrees.

Why so many definitions? It turns out that in a specific situation, one definition may be easier to use than the other. For example, suppose that you are given coordinates of $n$ points, and want to check whether the corresponding polygon is convex. The first definition requires $O(n^2)$ time, whereas the second definition leads to a linear-time algorithm.

Back to our problem, here the main difficulty is that the two polygons are in a general position. To specialize this problem, we need to somehow “align” them in a more convenient position. Figure 4.7 gives such a case, where all but one side $AB$ of $P$ doesn’t belong entirely to the perimeter of $Q$. In this case, all we need to do is to prove that the length of $AB$ is shorter than the corresponding curve from $A$ to $B$ in $Q$, but this follows from the triangle inequality.

Now, let’s get back to the general case. To make use of the special case here, we need somehow create a polygon $R$ that is well aligned with $Q$, as in the special case, and thus $\text{Perimeter}(R) < \text{Perimeter}(Q)$. This intermediate polygon $R$ somehow needs to be related to $P$, so that it’s easier to compare the perimeter of $R$ and that of $P$. A way to do that is to extend an edge of $P$ to cut the boundary of $Q$, as illustrated in Figure 4.8. An important observation here is that since $P$ is convex, by using the first definition of convexity, $P$ lies entirely inside $R$. (You should try an example with a concave $P$ to see why this argument fails in that case. This shows that the assumption that $P$ is convex is crucial.)

Now, we need to show that $\text{Perimeter}(P) \leq \text{Perimeter}(P)$. Let’s say that $P$ has $\ell$ edges. Note that here one edge of $P$ is aligned with an edge of $R$. We can recurse as above, finding another polygon $S$ that contains $P$. 

Figure 4.7: Polygons $P$ and $Q$ in a special position, where only one side of $P$ doesn't belong entirely to the perimeter of $Q$.

Figure 4.8: Polygons $P$ and $Q$ in a general position, with a polygon $R$ that is well aligned with $Q$ that contains $P$.

such that $\text{Perimeter}(S) < \text{Perimeter}(R)$, and two edges of $P$ are aligned with the edges of $S$. Eventually after at most $\ell$ steps, this process will terminate, and we’re done.

**Exercise 1:** We are given a square $ABCD$ with 4 points $M \in AB$, $N \in CD$, $P \in AD$, and $Q \in BC$, as illustrated in Figure 4.9. Suppose that $MN$ is perpendicular to $PQ$. Show that the segments $MN$ and $PQ$ have the same length.

**Exercise 2:** You are given a set of $n$ points with integer coordinates. Design an efficient algorithm to decide whether the set has a center of symmetry. A set of points $S$ has the center of symmetry if there exists a point $s$ (not necessarily in $S$) such that for every point $p \in S$ there is a point $q \in S$ such that $p - s = s - q$. See Figure 4.10 for an illustration.
4.4 The Superposition Pattern

4.4.1 Secret Sharing

How to distribute a secret. Suppose that we have a secret $X$ that is represented as a floating-point number. Let $n > t \geq 1$ be integers. We want to divide the secret to $n$ shares $S_1, \ldots, S_n$ such that (i) it’s easy to recover the secret given $t + 1$ shares, but (ii) it’s impossible to reconstruct the secret given just $t$ shares.

Let’s start with the simple case $n = 2$ and $t = 1$. A geometric way to solve the problem above is as follows. Let $A$ be the point $(0, X)$, and draw a random line passing through $A$. Let $B$ and $C$ be the points on the line with $x$-coordinates 1 and 2 respectively. Then $S_1$ and $S_2$ are the $y$-coordinates of $B$ and $C$, respectively. See Figure 4.11 for an illustration. Clearly, if we have both $S_1$ and $S_2$, we can reconstruct the line $BC$, find the point $A$, and recover $X$. If we have just one of the shares then we get one point in the line, but the secret can be anything.

The approach above continues to work for any choice of $n \geq 2$, as long as $t = 1$. Again, we pick a random line $f(x) = ax + b$ passing through $A = (0, X)$, meaning that $X = f(0) = b$. Then the share $S_i = f(i)$, for
Figure 4.11: Secret sharing, illustrated for $t = 1$ and $n = 2$.

every $i = 1, 2, \ldots, n$. If we have any two shares, we can reconstruct the line and recover $X$.

The key idea for the special case $t = 1$ is that, to determine a line $f(x) = ax + b$, we need exactly two points in the line, since there are two parameters $a$ and $b$. In general, if we have $t + 1$ points, we need $t + 1$ parameters, meaning that the corresponding function $f(x)$ should be a polynomial of degree $t$.

Specifically, for general $n$ and $t$, we will find a random polynomial $f(x)$ of degree $t$ that passes through the point $A = (0, X)$, that is, $X = f(0)$. Then $S_i = f(i)$, for every $i = 1, \ldots, n$. See Figure 4.12 for an illustration.

Figure 4.12: Secret sharing, illustrated for $t = 4$ and $n = 5$.

**THE INTERPOLATION PROBLEM.** What remains is to show how to recover the passing polynomial of degree $t$, given $t + 1$ points $(x_1, y_1), \ldots, (x_{t+1}, y_{t+1})$ in the plane. This is known as the interpolation problem.

Let’s start with the special case where all the points $y_i$ are 0. This is uninteresting, because the solution is trivial: $f(x) = 0$. To refine it to a more interesting special case, let’s instead only assume that $y_i = 0$ for every $i \neq 1$. Since $x_2, \ldots, x_{t+1}$ are the roots of the polynomial $f(x)$ of degree at most $t$, this polynomial must be of the form $C_1(x - x_2) \cdots (x - x_{t+1})$, for some constant $C_1$. To find the constant $C_1$, recall that $f(x_1) = y_1$. Thus

$$C_1 = \frac{y_1}{(x_1 - x_2)(x_1 - x_3) \cdots (x_1 - x_{t+1})}.$$
Now, in the example above, the abscissa $x_1$ plays a special role, distinct from the common role that falls to the other abscissas. Yet there is no peculiar virtue in the abscissa $x_1$: we can let any other given $x_r$ play that special role, and obtain the corresponding formulas $f_r(x) = C_r(x-x_1) \cdots (x-x_{r-1})(x-x_{r+1}) \cdots (x-x_{t+1})$.

We have now outlined the solution in $n$ different particular cases of our problem. Can we combine them so as to obtain the solution of the general case from the combination? Of course we can, by adding the $n$ special cases:

$$f(x) = f_1(x) + \cdots + f_{t+1}(x).$$

Note that for any $r \leq t + 1$, $f(x_r) = f_r(x_r)$ because $f_i(x_r) = 0$ for every $i \neq r$. Moreover, as $f_r(x_r) = y_r$, we have $f(x_r) = y_r$.

**Reflection.** In solving the interpolation problem, we start by finding accessible special cases, and then obtain the general solution by combining the special ones (via addition). Polya calls this the superposition pattern. Below is a toy exercise for you to practice this technique.

**Exercise:** Prove that the angle at the center of a circle is double the angle at the circumference on the same base, that is, on the same arc. See Figure 4.13 for an illustration.

![Figure 4.13: Three points A, B, C in a circle with center O. We want to show that $\angle BOC$ is twice $\angle BAC$.](image)

### 4.4.2 The Repertoire Method

While one can solve recurrences via the substitution method, its guessing trick is not reliable, as we are in the dark and have to fumble for a solution. Thus we need a more systematic way to find solutions. Let’s get back to $S_2(n) = 1^2 + 2^2 + \cdots + n^2$. First, note that we can express it as a recurrence $S_2(1) = 1$, and $S_2(n) = S_2(n-1) + n^2$ for every $n > 1$. Let’s now try to solve this recurrence via generalization and specialization.

**Generalization.** Here’s a way to generalize the problem. Fix real numbers $a, b, c$, and consider the recurrence $G_1 = 1$ and $G_n = G_{n-1} + an^2 + bn + c$ for every $n > 1$. Our problem is the special case $a = 1$ and $b = c = 0$. We’ll solve the general problem, and thus also obtain a solution for the original problem.

**Specialization:** Step 1. Now let’s get back to the generalized problem, and try to compute $G_n$ for some small values of $n$. We have

$$G_2 = G_1 + 4a + 2b + c = 1 + 4a + 2b + c.$$
Here we convert the summation form recurrence, so if you want to use it, you first need to convert your problem into a recurrence. For example, the repertoire method.

For our original problem, pick \(a\) namely \(G\)

Let's try to specialize the problem in a different way. Previously, we picked \(S\) and \(C\)

This suggests that we can use the superposition pattern here. In particular, if we can find the closed-formed formulas of \(G\) for three special cases, we can find \(A(n), B(n), C(n)\) and solve the general case.

**Specialization: Step 2.** Let’s first play with the problem by picking some specific \((a, b, c)\) that makes the problem easy to solve. Let’s start with \(a = b = 0\) and \(c = 1\). Thus our recurrence becomes: \(G_1 = 1\) and \(G_n = G_{n-1} + 1\) for every \(n > 1\). The solution is simple: \(G_n = n\). Next, let’s try \(a = 0, b = 1\) and \(c = 0\). Our recurrence becomes: \(G_1 = 1\) and \(G_n = G_{n-1} + n\) for every \(n > 1\). Thus \(G_n = 1 + \cdots + n\), and thus \(G_n = \frac{n(n+1)}{2}\).

**Specialization: Step 3.** Let’s try to specialize the problem in a different way. Previously, we picked \((a, b, c)\) and then computed \(G_n\). Let’s now reverse the process, by picking \(G_n\) and then computing \((a, b, c)\). But what specific formula for \(G_n\) should we pick? Our prior \(G_n\) formulas are \(G_n = n\) (a polynomial of degree 1), and \(G_n = \frac{n(n+1)}{2}\) (a polynomial of degree 2). So let’s try a polynomial of degree 3 this time, namely \(G_n = n^3\). Then

\[
a n^2 + bn + c = G_n - G_{n-1} = n^3 - (n-1)^3 = 3n^2 - 3n + 1.
\]

Hence \(a = 3, b = -3\) and \(c = 1\).

**Solving the general problem.** From the 3 steps of specialization above, we conclude that

\[
\begin{align*}
n & = 1 + 0 \cdot A(n) + 0 \cdot B(n) + 1 \cdot C(n) = 1 + C(n) \\
\frac{n(n+1)}{2} & = 1 + 0 \cdot A(n) + 1 \cdot B(n) + 0 \cdot C(n) = 1 + B(n) \\
n^3 & = 1 + 3 \cdot A(n) + (-3) \cdot B(n) + 1 \cdot C(n) = 1 + 3A(n) - 3B(n) + C(n) .
\end{align*}
\]

View the equations above as a system of linear equations with three variables \(A(n), B(n), C(n)\). Solving this system, we obtain \(C(n) = n - 1, B(n) = \frac{3n^2 - n - 2}{2}\), and \(A(n) = \frac{2n^3 + 3n^2 + n - 6}{6}\).

For our original problem, pick \(a = 1, b = 0, c = 0\) and thus

\[
S_2(n) = 1 + 1 \cdot A(n) + 0 \cdot B(n) + 0 \cdot C(n) = \frac{2n^3 + 3n^2 + n}{6}.
\]

**The repertoire method.** What you have just seen is called the repertoire method. It’s a way to solve recurrence, so if you want to use it, you first need to convert your problem into a recurrence. For example, here we convert the summation form \(G_n = 1^2 + \cdots + n^2\) to a recurrence: \(G_1 = 1\) and \(G_n = G_{n-1} + n^2\). The repertoire method consists of the following steps:

1. Generalize the problem into a parameterized one. For example, here we have a generalized problem: \(G_1 = 1\) and \(G_n = G_{n-1} + an^2 + bn + c\).
2. Find a way to represent solutions of the general problem via the parameters. For example, we realize that \(G_n\) must be of the form \(1 + A(n) \cdot a + B(n) \cdot b + C(n) \cdot c\).
3. Solve the problem for some special cases. This can work both ways. For example you can pick \((a, b, c) = (0, 0, 1)\) to find \(G_n = n\), or conversely, you can pick \(G_n = n^3\) to find \((a, b, c) = (3, -3, 1)\). If you have \(k\) parameters (here \(k = 3\), since we have three parameters \(A, B, C\)) then you need to consider \(k\) well-chosen special cases.

4. Plug the parameterized solution to the special cases and solve a system of equations. Here after solving the system of equations, you obtain \(C(n) = n - 1, B(n) = \frac{n^2 + n - 2}{2}\) and \(A(n) = \frac{2n^3 + 3n^2 + n - 6}{6}\).

**Exercise 1:** Let \(F_0 = 1\) and \(F_1 = 1\), and \(F_n = F_{n-1} + F_{n-2}\) for every \(n \geq 3\). This is the Fibonacci sequence. Use the repertoire method, find the closed-form formula of \(F_n\).

**Exercise 2:** During the Jewish-Roman war, Josephus was among \(n\) Jewish rebels that were trapped in a cave by the Romans. Preferring suicide to capture, the rebels decided to form a circle, and proceeding around it, to kill every second remaining person until only one survives. Josephus didn’t want such suicide nonsense, so he needed to calculate the surviving place \(J(n)\) in the circle.

For example, consider \(n = 10\). After the first round, people at even positions (namely 2, 4, 6, 8, 10) will be eliminated. In the subsequent rounds, 3, 7, 1, 9 will be eliminated in that order, and 5 will be the survivor. Thus \(J(10) = 5\).

Use the repertoire method, find a closed-form formula for \(J(n)\).