Bachelorarbeit

Andreas Windorfer

3. April 2020

Zusammenfassung

Inhaltsverzeichnis

1 Einleitung

2 Fazit

3 Dynamische binäre Suchbäume

innerer Knoten
Teilbaum
Höhe Nur eine Wurzel vorhanden, dann Höhe 0
innerer Knoten, auch Wurzel
Rotationen
Einfügen
Suchen
Löschen
Bruder
Höhe
Pfad

4 Balancierte Suchbäume

4.1 Rot-Schwarz-Baum

Der Rot-Schwarz-Baum gehört zur Gruppe der balancierten binären Suchbäume. Zusätzlich zu den Zeigern *links*, *rechts* und *vorgaenger* benötigt jeder Knoten ein zusätzliches Bit Speicherplatz, um die Farbinformation zu speichern. Der Name der Datenstruktur kommt daher, dass die beiden durch das zusätzliche Bit unterschiedenen Zustände als *Rot* und *Schwarz* bezeichnet werden. Die Farbe ist also eine Eigenschaft der Knoten und im folgenden wird einfach von roten bzw. schwarzen Knoten gesprochen. Zeiger auf Nachfolger oder Vorgänger die aktuell im Baum nicht vorhanden sind zeigen auf einen schwarzen Sonderknoten. Dieser Sonderknoten enthält in dieser Arbeit den Schlüsselwert *null*.
Folgende zusätzliche Eigenschaften müssen bei einem Rot-Schwarz-Baum erfüllt sein.

1. Jeder Knoten ist entweder rot oder schwarz.

2. Die Wurzel ist schwarz.

3. Jedes Blatt ist schwarz.

4. Beide Nachfolger eines roten Knotens sind schwarz.

5. Für jeden Knoten enthalten alle Pfade, die an diesem Knoten starten und in einem Blatt enden, die gleiche Anzahl an schwarzen Knoten.

Der Zweck dieser Einschränkungen ist es die Höhe des Rot-Schwarz-Baumes zu begrenzen, um die üblichen Operationen effizient ausführen zu können. Als **Schwarz-Höhe** $bh(x)$ eines Knoten x wird die Anzahl der schwarzen Knoten in einem Pfad, der an dem Knoten startet und in einem Blatt endet, bezeichnet. Die eigene Farbe des betrachteten Knotens bleibt dabei außen vor. Aufgrund der Eigenschaft fünf ist die Schwarz-Höhe eines Knotens wohldefiniert. Die Höhe h ist beim Rot-Schwarz-Baum wie bei anderen binären Suchbäumen definiert und berücksichtigt auch die roten Knoten. Die Schwarz-Höhe der Wurzel des Baumes in Abbildung 1 ist also zwei. Im Folgenden wird manchmal nur die Nummer einer Eigenschaft verwendet. Bezieht sich der Text also beispielsweise auf Eigenschaft zwei, ist damit die Eigenschaft bzgl. der schwarzen Wurzel gemeint.

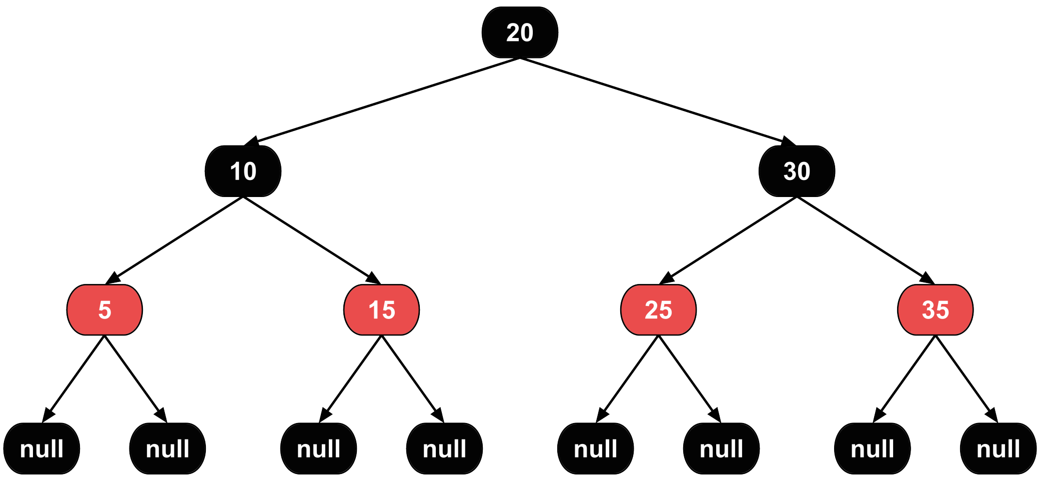

Abbildung 1: Rot-Schwarz-Baum ohne Verletzung von Eigenschaften.

4.1.1 Einfügen in den Rot-Schwarz-Baum

Ein neuer Schlüssel wird zunächst wie in Ref: (normales Einfügen) eingefügt. Zusätzlich werden dann noch die Zeiger *links* und *rechts* auf den schwarzen Sonderknoten gesetzt. Die Farbe neu eingefügter Knoten ist rot. Durch den neu eingefügten Knoten können Korrekturen notwendig werden, um die Rot-Schwarz-Baum Eigenschaften zu erhalten. Zunächst betrachten wir der

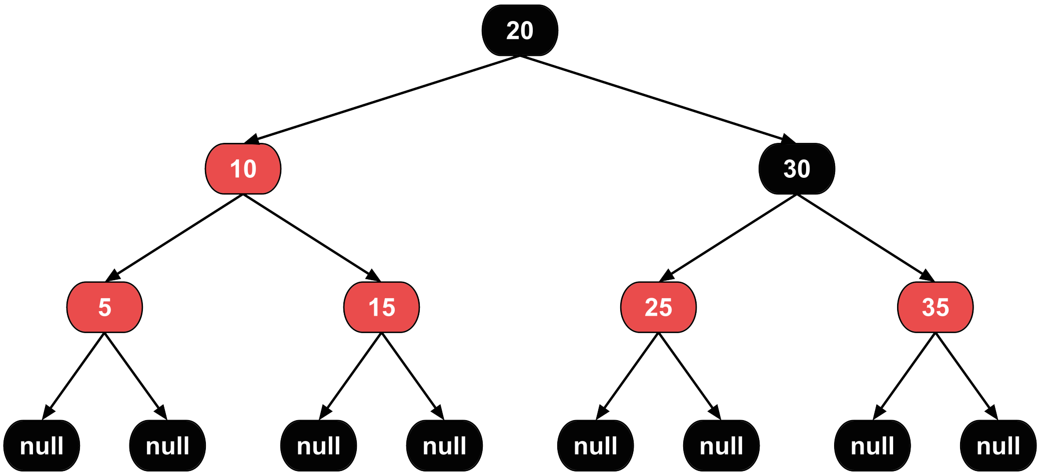

Abbildung 2: Rot-Schwarz-Baum bei dem Eigenschaft vier und fünf verletzt sind.

Reihe nach welche der fünf Eigenschaften betroffen sein können. Es wird davon ausgegangen, dass in einen Rot-Schwarz-Baum ohne Verletzung von Eigenschaften eingefügt wurde.

1. Es ist immer noch jeder Knoten entweder rot oder schwarz.
2. Wurde in den leeren Baum eingefügt, so ist der neu eingefügte rote Knoten die Wurzel, was eine Verletzung darstellt. Waren bereits Knoten im Baum vorhanden bleibt die Wurzel unverändert.
3. Aufgrund des Sonderknotens sind die Blätter immer noch schwarz.
4. Der Baum wird nur direkt an der Einfügestelle verändert. Der neue Knoten hat schwarze Nachfolger, er könnte jedoch einen roten Vorgänger haben, so dass diese Eigenschaft verletzt wäre.
5. Da der neue Knoten rot ist, ändern sich keine Schwarz-Höhen von bereits enthaltenen Knoten. Die Schwarz-Höhe des neuen Knoten ist immer 1, denn beide Nachfolger verweisen auf den schwarzen Sonderknoten. Eigenschaft fünf bleibt also erhalten.

Es können also die Eigenschaften zwei und vier betroffen sein. Jedoch nur eine von ihnen, denn Eigenschaft zwei wird genau dann verletzt wenn der neue Knoten der Einzige im Baum ist. Dann kann er aber keinen roten Vorgänger haben.
Zur Korrektur wird eine zusätzliche Routine eingesetzt. Diese Routine arbeitet sich von der Einfügestelle solange nach oben in einer Schleife durch, bis

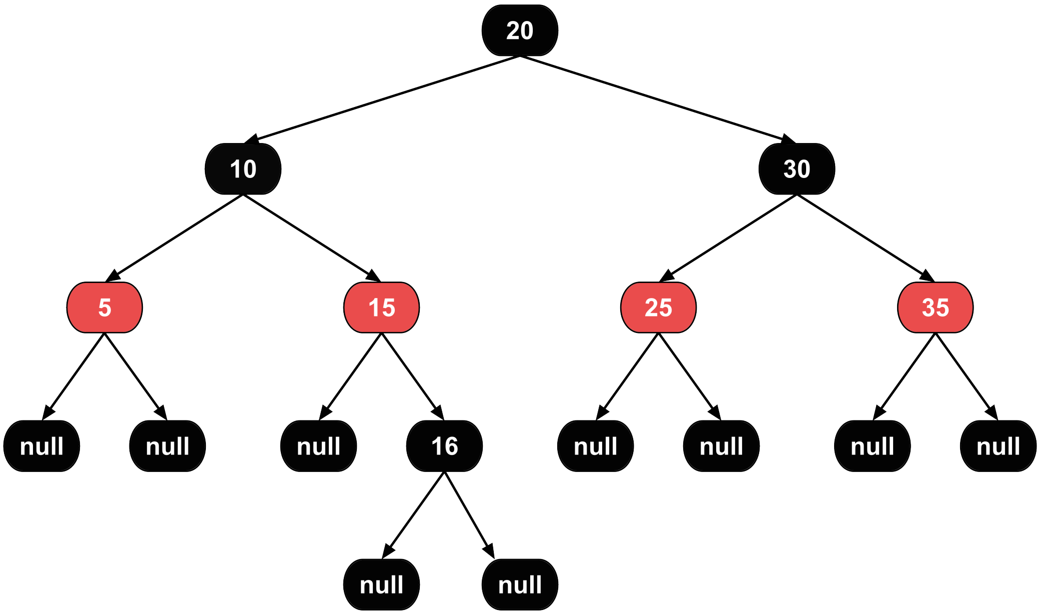

Abbildung 3: Rot-Schwarz-Baum bei dem Eigenschaft fünf verletzt ist.

alle Eigenschaften wieder erfüllt sind. Die Schleifenbedingung ist, dass eine Verletzung vorliegt. Dazu muss geprüft werden ob der betrachtete Knoten die rote Wurzel des Gesamtbaumes ist, oder ob er und sein Vorgänger beide rot sind. Beim ersten Durchlauf wird der neu eingefügte Knoten übergeben. Innerhalb der Schleife werden sechs Fälle unterschieden. Im folgenden wird auf vier Fälle detailliert eingegangen. Die restlichen zwei verhalten sich symmetrisch zu einem solchen. Jeder der Fälle verantwortet, dass zum Start der nächsten Iteration wieder nur maximal eine der beiden Eigenschaften zwei oder vier verletzt sein kann und Eigenschaft vier lediglich an einer Stelle verletzt wird. Eigenschaft eins bleibt in der Beschreibung außen vor, da es während der gesamten Laufzeit der Routine nur Knoten gibt, die entweder rot oder schwarz sind.

Fall 1: Die Wurzel ist rot: Dieser Fall wird behandelt in dem man die Wurzel schwarz färbt. Man muss noch zeigen, dass es durch das Umfärben zu keiner anderen Verletzung gekommen ist.

1. -
2. Die Wurzel wurde schwarz gefärbt.
3. Die Blätter sind unverändert.

4. Es wurden weder rote Knoten hinzugefügt, noch wurden Zeiger umgesetzt.

5. Da Umfärben der Wurzel kann hierauf keinen Einfluss haben, da sie in der Berechnung der Schwarz-Höhe jedes Knotens außen vor ist.

Es wird also keine Eigenschaft mehr verletzt und die Schleife wird keine weitere Iteration durchführen.
Die Fälle 2 - 6 behandeln nun die Situation zweier aufeinanderfolgender roter Knoten. Der untere rote Knoten wird als x bezeichnet, der obere als y. Da Eigenschaft fünf nach jeder Iteration erfüllt ist muss y einen Bruder haben. Denn da y rot ist und Fall 1 nicht ausgewählt wurde, kann es nicht die Wurzel sein. Also muss auch y einen Vorgänger z haben. Von z aus starten über y mindestens zwei Pfade zu schwarzen Blättern. Hätte z nicht auch einen rechten Nachfolger, wäre bereits vor der Iteration Eigenschaft fünf verletzt gewesen. Wenn sich auf die Situation nach der Fallbehandlung bezogen wird, wird den Variablen ein Hochstrich hinzugefügt. Wurde beispielsweise Knoten y von rot nach schwarz gefärbt, so ist y rot und y' schwarz.
Fall 2: Der Bruder von y ist rot: Diesen Fall veranschaulicht Abbildung 4. Da y rot ist, muss z nach Eigenschaft schwarz sein, ansonsten wäre Eigenschaft vier mehrfach verletzt gewesen. Nun wird z rot gefärbt und beide Nachfolger von z, also y und dessen Bruder, schwarz. Somit hat x nun einen schwarzen Vorgänger und die Verletzung der Eigenschaft vier wurde an dieser Stelle behoben. Wie sieht es aber mit den Verletzungen insgesamt aus ? Der rot umgefärbte Knoten z hat zwei Nachfolger, somit wurde kein Blatt rot gefärbt. Die Schwarz-Höhen der Pfade mit z als inneren Knoten ändern sich nicht, da sie auch genau eines der schwarz umgefärbten Nachfolger von z enthalten müssen. Die Schwarz-Höhen der Pfade mit z als Startknoten werden um eins verringert. Dies gilt jedoch für jeden von ihnen, so dass Eigenschaft fünf weiterhin nicht verletzt wird. Wenn der rot gefärbte Knoten z nicht die Wurzel ist, könnte er einen roten Vorgänger haben und Eigenschaft vier ist weiterhin Verletzt, das Problem liegt nun aber zwei Baumebenen höher. Wenn z die Wurzel ist kann Eigenschaft vier nicht mehr verletzt sein, allerdings hat man nun eine rote Wurzel. Man ist also in der Situation die nächste Iteration durchführen zu können.

Fall 3: y ist ein linker Nachfolger und sein Bruder ist schwarz. x ist linker Nachfolger:
Abbildung 5 zeigt eine entsprechende Situation. Da in dieser Situation, die Wurzel schwarz sein muss, gibt es den Vorgänger z von y. Da es nur eine Stelle im Baum geben kann an der Eigenschaft vier verletzt wird, muss z

schwarz sein. Es wird nun eine Rechtsrotation auf z ausgeführt. Anschließend wird z rot gefärbt und y schwarz. Die Eigenschaften 2 - 5 werden nun wieder nacheinander betrachtet. Dazu werden vier weitere Variablen auf Knoten verwendet. Es zeigt xl den linken Nachfolger von x, xr entsprechend den rechten Nachfolger. yr und zr bezeichnen die rechten Nachfolger von y bzw. z. Nachfolgend wird verwendet, dass die Teilbäume mit Wurzel xl, xr, yr und zr durch die Ausführung unverändert bleiben. Wenn z zu Beginn nicht die Wurzel des Gesamtbaumes war, blieb diese unbearbeitet. Ansonsten ist wurde durch die Rotation y zur neuen Wurzel und y wurde schwarz gefärbt. Alle vier Plätze in der zweiten Ebene unter z werden von den unveränderten Teilbäumen mit den Wurzeln xl, xr, yr oder zr besetzt. An den Blättern verändert sich also durch die Ausführung nichts.
Nun wird betrachtet ob es noch aufeinanderfolgende rote Knoten im Baum gibt. Knoten x ist jetzt direkter linker Nachfolger der neuen schwarzen Wurzel y des Teilbaumes. An den nachfolgenden Teilbäumen von x hat sich nichts geändert. Die linke Seite des neu entstandenen Baumes enthält also keine aufeinanderfolgenden roten Knoten. Der rechte Nachfolger von y ist nun das rote z. Rechts an z hängt nun eine unveränderter Teilbaum, dessen Wurzel zuvor Bruder von y war. Dieser ist nach Fallunterscheidung ein schwarzer Knoten. Links hängt ebenfalls ein unveränderter Teilbaum, dessen Wurzel zuvor rechter Nachfolger von y war. Der rechts Nachfolger von y muss zu Beginn schwarz gewesen sein, ansonsten wäre Eigenschaft vier an zwei Stellen gegeben, die Eigenschaft vier verletzten. Im Teilbaum mit Wurzel y gibt es also keine aufeinanderfolgenden roten Knoten. Da y schwarz gefärbt wurde kann auch nach oben keine neue Verletzung entstanden sein. Wir kommen zu den Schwarz-Höhen und betrachten zunächst die Ausgangssituation. Es gilt $bh(xl) = bh(xr) = bh(yr) = bh(zr) = bh(z) - 1$. Nun betrachten wir den Baum nach der Änderung. Wie oben bereits erwähnt wird die zweite Ebene unter der Wurzel von den Knoten xl', xr', yr' und zr' gebildet. Es müssen also lediglich die Knoten x', y' und z' betrachtet werden. An x' und an z' folgen schwarze Knoten mit der Schwarz-Höhe $bh(z) - 1$. Die Schwarz-Höhen von x' und z' ist also wohldefiniert und es gilt $mathitbh(x') = bh(z') = bh(z)$. y' Nachfolger sind die roten Knoten x' und z'. Da die beide Nachfolger rot sind gilt $bh(y') = mathitbh(x') = bh(z')$. Somit sind alle Schwarz-Höhen im betrachteten Teilbaum in Ordnung. Die neue Wurzel der Teilbaumes y' hat die gleiche Schwarz-Höhe und die gleiche Farbe wie die vorherige Wurzel z. Damit kann es auch im Gesamtbaum zu keiner Verletzung der Eigenschaft gekommen sein. Es ist keine der Eigenschaften verletzt, daher wird es zu keiner Iteration mehr kommen. **Fall 4: y ist ein linker Nachfolger und sein Bruder ist schwarz. x ist rechter Nachfolger:**
Dieser in Abbildung 6 gezeigte Fall wird so umgeformt, dass eine Situation

entsteht bei der Fall drei angewendet werden kann. Dazu wird eine Linksrotation an Knoten y durchgeführt. Wieder müssen die Auswirkungen auf die Eigenschaften 2- 5 betrachtet werden. Veränderungen gab es durch die Rotation lediglich im Teilbaum mit Wurzel z.Ist z die Wurzel des Gesamtbaumes, so ist es auch z' und z' ist ein schwarzer Knoten. Es seien xl und xr die linken bzw. schwarzen Nachfolgeknoten von x. yl ist der linke Nachfolger von y. Die Teilbäume xl, xr und yr enthalten alle Blätter innerhalb des linken Teilbaumes von z. Außerdem bleiben die Teilbäume xl, xr und yr durch die Rotation unverändert und sie enthalten auch alle Blätter des linken Teilbaumes von z'. Der rechte Teilbaum von z bleibt unverändert. Eigenschaft drei wird also weiterhin eingehalten.

Da x und y rot sind müssen z, xl und xr schwarz sein. Durch die Rotation ist y' linker Nachfolger von x'. x' ist Nachfolger vom schwarzen z'. Alle verbleibenden Nachfolger von x' und y' werden durch die unveränderten Teilbäume xl, xr und yr gebildet. Deren Wurzeln müssen schwarz sein, ansonsten hätte es in ursprünglichen Baum an mehr als einer Stelle eine Verletzung von Eigenschaft vier gegeben. Durch die Rotation verbleibt es also bei einer Verletzung der Eigenschaft vier in der gleiche Baumebene. Die beiden beteiligten roten Knoten hängen nun aber links nun aber links.

4.1.2 Löschen aus dem Rot-Schwarz-Baum

4.1.3 Suchen im Rot-Schwarz-Baum

Lemma: Maximale Höhe des Rot-Schwarz-Baum:
Für einen Rot-Schwarz-Baum mit Höhe h und n inneren Knoten gilt $h \leq 2lg(n+1)$.
Beweis:
Zunächst wird gezeigt, dass ein Teilbaum t_1 mit Wurzel x_1 und Schwarzhöhe $bh(x_1)$ zumindest über $2^{bh(x_1)} - 1$ innere Knoten verfügt, also $n \geq 2^{bh(x_1)} - 1$. Dies wird durch Induktion über die Höhe h_{x_1} gezeigt. Für $h_{x_1} = 0$ besteht t_1 nur aus einem Blatt und enthält keine inneren Knoten. Natürlich gilt in diesem Fall auch $bh(t_1) = 0$.
Induktionsanfang mit $h_{x_1} = 0$:
$2^0 - 1 = 0$
Induktionsschritt:
Nun wird ein Teilbaum T_2 mit Wurzel x_2 mit Höhe $h+1$ betrachtet. Jedes seiner beiden Nachfolger hat entweder Schwarzhöhe $bh(x_2)$, wenn es rot ist, oder $bh(x_2) - 1$, wenn es schwarz ist. Die Höhe beider Nachfolger ist niedriger als die eigene Höhe von x_2. Somit kann bei beiden Nachfolgern jeweils die Induktionsnahme für die Mindestanzahl der inneren Knoten eingesetzt

werden.
$2^{bh(x_1)-1} - 1 + 2^{bh(x_1)-1} - 1 = 2^{bh(x_1)} - 2$
Addiert man einen inneren Knoten aufgrund der Wurzel x_2 hinzu, erhält man die Behauptung.
$2^{bh(x_1)} - 2 + 1 = 2^{bh(x_1)} - 1.$
Es gilt also $n \geq 2^{bh(x)} - 1$.
Auf einem Pfad in einem Rot-Schwarz-Baum t_3 von der Wurzel bis zu einem Blatt sind der erste und der letzte Knoten, sowie mindestens jeder zweite der weiteren Knoten schwarz. Es gilt also $bh(t_3) \geq \frac{h(t_3)}{2}$. Damit kann man in der Ungleichung $bh(t_3)$ durch $h(t_3)$ ersetzen, woraus dann das Lemma folgt.
$n \geq 2^{\frac{h(t_3)}{2}} - 1 \Rightarrow n + 1 \geq 2^{\frac{h(t_3)}{2}} \Rightarrow \lg(n+1) \geq \frac{h(t_3)}{2} \Rightarrow 2lg(n+1) \geq h(t_3)$

Literatur

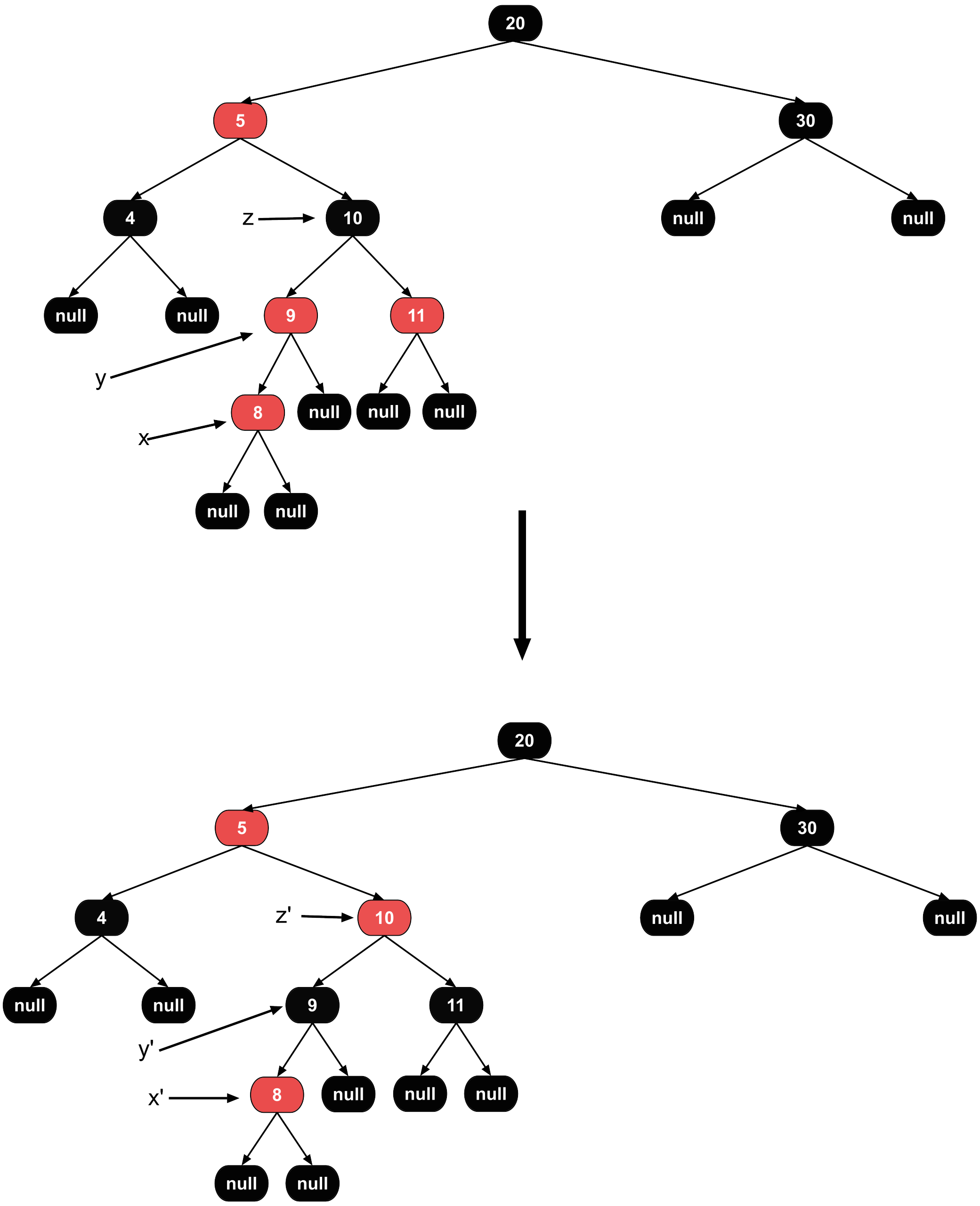

Abbildung 4: Korrektur nach dem Einfügen. Dargestellt ist Fall 2

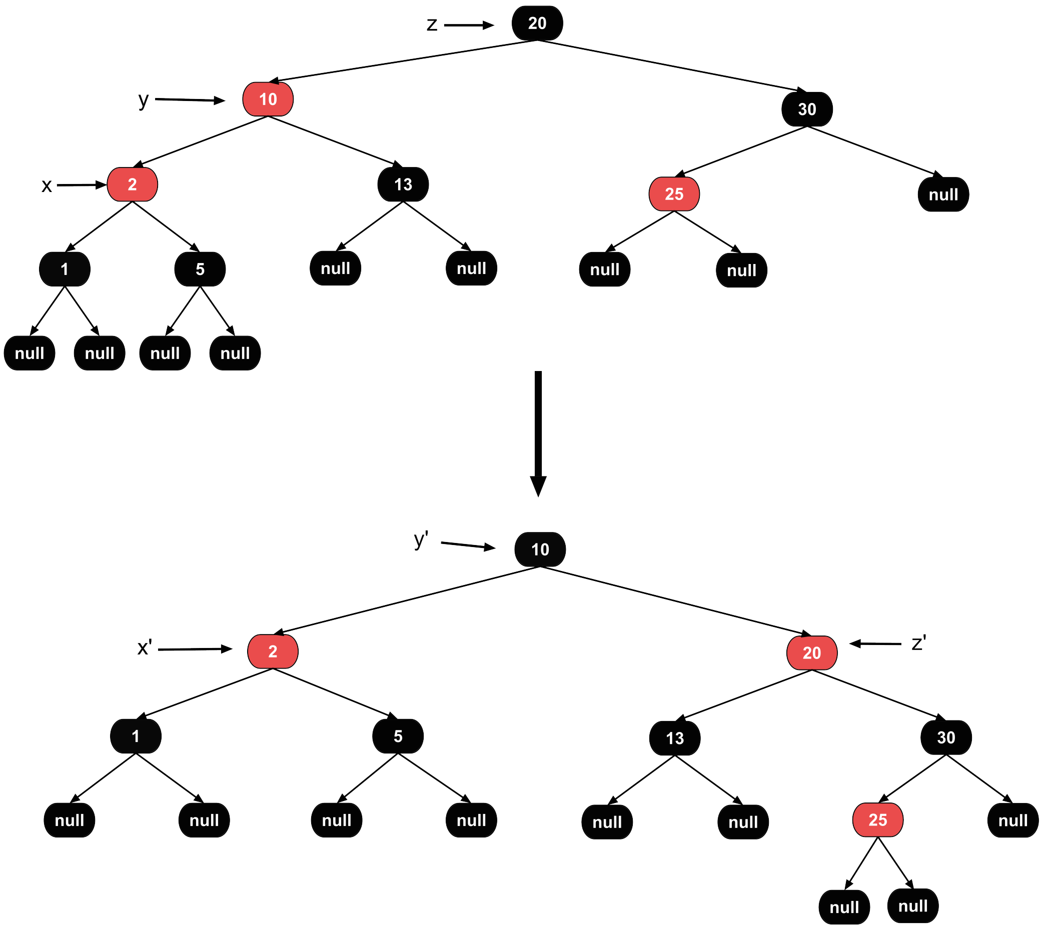

Abbildung 5: Korrektur nach dem Einfügen. Dargestellt ist Fall 3

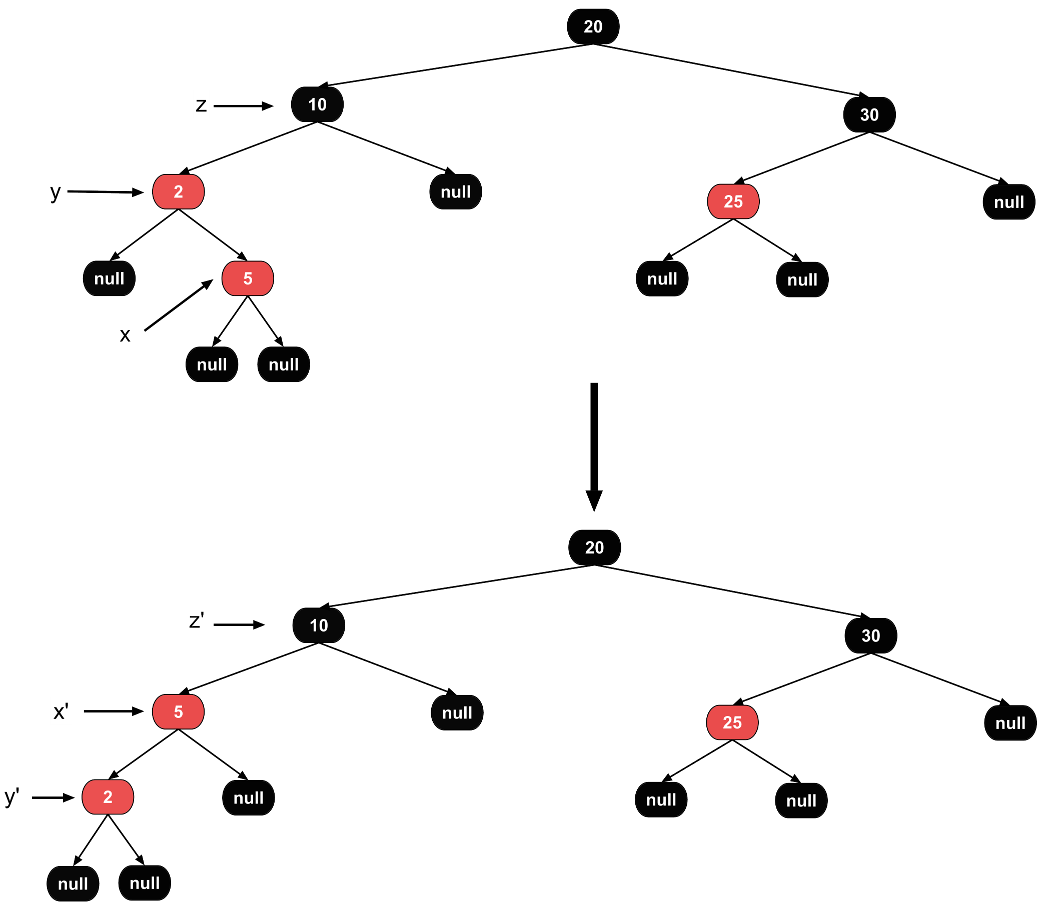

Abbildung 6: Korrektur nach dem Einfügen. Dargestellt ist Fall 4