TP 11
Dichotomie

Le mot **dichotomie** provient du grec et signifie « couper en deux ». En informatique, il s'agit ici d'une méthode générale de recherche de solution d'un problème en coupant l'intervalle de recherche en deux à chaque étape. Nous allons d'abord l'illustrer par un jeu simple et bien connu.

I Préliminaire : un jeu

On connait le jeu suivant entre deux joueurs : l'un pense à un nombre entre 1 et 100 et l'autre doit le deviner en formulant ses propositions, auxquelles la réponse sera seulement « plus petit », « plus grand » ou bien « trouvé ! ».

Il est tout à fait possible de programmer ce jeu, où c'est le programme qui choisit le nombre au hasard !

Exercice 1
Compléter le programme fourni : il choisit un nombre au hasard, l'affiche (au début, pour faire des tests, et ensuite on supprimera la ligne) puis demande à l'utilisateur une proposition de nombre. Il dit ensuite si le nombre à trouver est plus grand, plus petit, ou si c'est bon.

Jouez à ce jeu pendant quelques minutes.

Exercice 2
Modifier le programme pour qu'en plus il compte le nombre de tentatives du joueur, affichant à la fin un « trouvé en n coups », et tenter de gagner en un minimum de coups possibles.

On peut aussi apporter diverses modifications comme par exemple choisir les nombres dans un intervalle plus grand, ce qui augmente la difficulté.

Exercice 3
Quelle semble être la meilleure stratégie possible, en rapport avec le titre de ce TP ? Pour un nombre entre 1 et 100, en combien de coups êtes-vous absolument certain de gagner ? Plus généralement pour un nombre entre 1 et N, pouvez donner une formule pour le nombre de coups maximal en lesquels on est certain de gagner, et le démontrer ?

II Application aux solutions d'équations

On s'inspire de cette méthode pour trouver les solutions d'une équation. Soit une fonction f définie sur un intervalle $[a, b]$ à valeurs dans $\mathbb{R}$, continue et croissante, on cherche une solution à l'équation $f(x) = y$ en supposant $f(a) < y < f(b)$. Alors on calcule $m = \frac{a+b}{2}$ et on compare $f(m)$ avec y. Si $f(m) < y$ on doit chercher x entre m et b, et si $f(m) > y$ on doit chercher x entre a et m. Cela divise par deux la taille de l'intervalle de recherche.

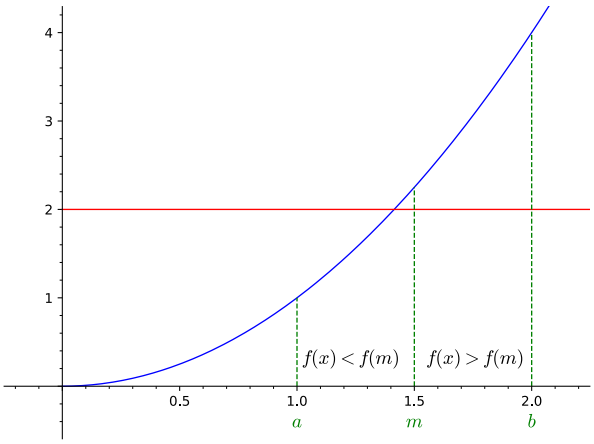

Figure 1. – Recherche par dichotomie de la solution à $f(x) = 2$, où $f : x \mapsto x^2$.

Remarque. On pourra traduire ces idées en une *preuve* du théorème des valeurs intermédiaires. Mais cela nécessite d'abord quelques pré-requis sur la continuité…

Exercice 4 *Échauffement*
On cherche les approximations successives de $\sqrt{2}$ qu'on obtiendrait avec cette méthode (mais sans utiliser la fonction racine carrée). Ici c'est la fonction $f : x \mapsto x^2$, et on sait $f(1) < 2 < f(2)$, autrement dit $1 < \sqrt{2} < 2$. À l'étape suivante on calculera $m = \frac{1+2}{2} = \frac{3}{2}$ puis $m^2 = \frac{9}{4} > 2$. Ceci *démontre* que $1 < \sqrt{2} < \frac{3}{2}$ (de l'autre côté, si $\frac{3}{2} < x < 2$ alors $x^2 > 2$). On peut continuer en prenant le milieu entre 1 et $\frac{3}{2}$: c'est $m = \frac{1}{2}\left(1 + \frac{3}{2}\right) = \frac{5}{4}$, et on calcule $m^2 = \frac{25}{16}$, mais $m^2 < 2 = \frac{32}{16}$, donc ceci démontre $\frac{5}{4} < \sqrt{2} < \frac{3}{2}$ soit $1{,}25 < \sqrt{2} < 1{,}5$. Et recommencer, etc.

1. Écrire une fonction `racine(n)` qui effectue n étapes de ce procédé, en affichant à chaque fois les bornes (a, b) de l'encadrement obtenu.
2. Améliorer la fonction en `racine2(s)` qui prend en argument un seuil $s > 0$ (on lui donnera par exemple $s = 10^{-6}$ avec la syntaxe `racine2(1e-6)`), et s'arrête quand l'encadrement (a, b) vérifie $b - a < s$. Si tout se passe bien on a à tout moment $a < b$ dans la boucle.
3. Bonus : la dichotomie est naturellement un algorithme récursif, car chercher par dichotomie la solution entre a et b revient à chercher par dichotomie la solution soit entre a et m soit entre m et b. Il faut alors mettre les bornes de l'intervalle en tant qu'arguments de la fonction. Écrire la fonction récursive `racine3(s, a, b)` qui renvoie le couple (a, b) tel que la solution est garantie être dans (a, b) avec $b - a < s$.

Exercice 5
Soit la fonction $f : x \mapsto x^3 - 5x - 3$.

1. Avec un tableau de variations, montrer que l'équation $f(x) = 0$ admet exactement trois solutions x_1, x_2, x_3. En calculant quelques valeurs particulières de f, donner un encadrement (par des nombres entiers) de chacune de ces solutions.
2. En déduire trois fonctions `solution1(s)`, `solution2(s)`, `solution3(s)`, donnant chacune un encadrement de la solution avec un écart strictement inférieur à $s > 0$. Attention car f est décroissante autour de x_2 !

Exercice 6
On s'intéresse maintenant à l'équation $(E_t) : e^x = 3 + tx$, d'inconnue $x \in \mathbb{R}$, mais avec un paramètre réel $t > 0$. On pourra utiliser la fonction `exp` du module `math` mais dans tous les cas il n'existe pas de formule simple pour la solution de cette équation. On pose $f_t : x \mapsto e^x - 3 - tx$.

1. Avec un tableau de variations (et un petit calcul de limites), justifier que l'équation $f_t(x) = 0$ admet bien deux solutions, une $x_t > 0$ et une autre $y_t < 0$.

Diverses intuitions ou représentations graphiques permettent de se rendre compte de la chose suivante : plus t est grand plus la solution x_t part vers $+\infty$; et plus t est petit plus la solution y_t part vers $-\infty$. On ne peut donc pas raisonnablement donner un encadrement de x_t et de y_t a priori qui soit indépendant de t.

2. Écrire une fonction `borne_x(t)` qui prend en argument le paramètre $t > 0$ et qui renvoie le *plus petit entier* $n \geqslant 0$ tel que $f_t(n) \geqslant 0$.
3. En déduire une fonction `solution_x(t, s)` qui prend en argument le paramètre $t > 0$ et un seuil $s > 0$, et cherche par dichotomie la solution x_t avec un seuil s, ou au départ x_t est supposé être entre 0 et la borne renvoyée par la fonction précédente.
4. De même, écrire une fonction `borne_y(t)` qui renvoie le plus petit entier $n \leq 0$ tel que $f_t(n) \geqslant 0$, puis une fonction `solution_y(t, s)` qui cherche par dichotomie la solution y_t avec le seuil s.

III Application à la recherche dans une liste

La dichotomie est une méthode terriblement efficace pour rechercher un élément dans une liste à condition qu'elle soit **triée**.

Exercice 7
Révisions : écrire une fonction `est_croissante(L)` qui renvoie `True` si la liste `L` est triée, et `False` sinon.

Exercice 8
Révisions : écrire une fonction `cherche_iteratif(L, x)` qui parcourt la liste et donne le premier indice où l'élément `x` apparait, et renvoie `None` s'il n'apparait pas.

La méthode par dichotomie pour chercher à quel indice apparait l'élément `x` consiste à partir d'une liste triée `L` de longueur n et à comparer l'élément `x` avec un élément au milieu de `L`. Si `x` est plus grand, c'est qu'il faut chercher dans la moitié de droite de la liste, ceux qui sont plus grand que le milieu, et sinon il faut chercher dans la moitié de gauche. On travaillera donc en initialisant deux variables : `a` à 0 et `b` à $n-1$ (le dernier indice de la liste).

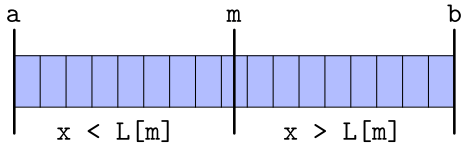

Figure 2. – Dichotomie dans une liste.

Par rapport au cas des fonctions continues, les nouveaux problèmes suivants se posent :

1. Il n'y a pas toujours exactement un élément au milieu de la liste (cela dépend de la parité de n), on ne peut donc pas à n'importe quel moment comparer `x` avec `L[(a+b)/2]`. D'une part `L[3.5]` affichera une erreur, mais en fait même `L[3.0]`, il faut donc travailler avec la division en nombres entiers `(a+b) // 2`. Cela revient à comparer `x` avec l'élément de `L` au milieu dans la moitié gauche si n est pair ; par exemple dans une liste de longueur 8 le dernier indice est 7, la première moitié correspond aux indices de 0 à 3 et la deuxième moitié de 4 à 7, et `(0+7) // 2 = 3` donc on comparera `x` avec `L[3]`.
2. L'élément `x` n'est pas nécessairement dans la liste, et on ne peut pas diviser par deux les intervalles à l'infini. Il faut donc savoir s'arrêter quand l'écart entre a et b devient exactement 1 et détecter alors si on trouve l'élément `x` à l'indice a ou b ou pas du tout.
3. Dans le cas où `x` serait strictement plus grand que tous les éléments de la liste, l'algorithme décrit ici s'appliquerait jusqu'au bout pour faire remonter la borne a jusqu'à b ; de même si `x` était plus petit que le minimum, l'algorithme décrit descendrait jusqu'à a. Il est donc préférable de tester dès le départ si `x` est bien compris entre `L[0]` et `L[n-1]`.

Exercice 9
Écrire la fonction `cherche_dichotomie(L, x)` qui recherche par cette méthode de dichotomie si l'élément `x` est dans la liste `L`, et renvoie son indice si elle le trouve et `None` s'il n'est pas dans la liste.

Les fichiers ci-joints contiennent un dictionnaire français de plus de 600 000 mots issu du logiciel libre de correction orthographique GNU Aspell, et un bout de programme qui va charger le dictionnaire comme une liste de mots. Sur les mots, l'opération Python `<` correspond à l'ordre lexicographique, ainsi la fonction précédente s'applique directement si `L` est la variable nommée `dictionnaire` et si `x` est un mot. Si tout se passe bien, le dictionnaire a déjà été trié dans l'ordre de Python. Cela provoque quelques étrangetés pour un humain : les mots avec des majuscules sont rangés d'abord, les mots commençant par des lettres accentuées sont rangés à la fin. Peu importe cependant, tant que le dictionnaire est bien rangé par ordre croissant *au sens* de l'opération `<` telle qu'elle est construite en Python.

Exercice 10
Avec le fichier joint, tester la recherche dans le dictionnaire, à la fois itératif et par dichotomie. On pourra modifier la fonction `cherche_dichotomie(L, x)` pour afficher le nombre d'étapes qu'elle réalise, et comparer cette valeur avec le logarithme en base 2 de la longueur du dictionnaire. Éventuellement renommer `cherche_dichotomie_print` la version qui affiche le nombre de coups.

Exercice 11
Le fichier joint contient un morceau de programme pour chronométrer le code Python, avec la fonction `timeit` du module `timeit`. Elle prend en argument le nombre de répétitions qu'elle va effectuer, et le résultat est exprimé en secondes. Tester le chronométrage, sur divers mots, à la fois avec les recherches itératives et dichotomiques (attention, pour chronométrer sa fonction il est important qu'elle ne fasse pas de `print`), avec au maximum un millier de répétitions.