26 mai 2025
Master 1 M.F., Orsay

Corrigé du Partiel de M.A.O. Calcul Formel Durée : 3 heures

Ce sujet est consacré au test de primalité de Miller-Rabin.

Partie 1 : Notion de témoin

1. ($\star$) Voir le Notebook ! On obtient $s = 7$ pour $n = 4225$.
2. Soit $a \in [\![2, n-1]\!]$ tel que $\text{pgcd}(a, n) \neq 1$. Alors pour tout entier $j \geq 1$, on a $a^j \not\equiv \pm 1 \bmod n$ sinon a serait inversible modulo n, d'inverse $\pm a^{j-1}$. Donc a est un témoin de n.
3. Soit $a \in [\![2, n-1]\!]$ un témoin de n, avec $n = p$ premier. Considérons le plus petit entier $i \in [\![0, s]\!]$ tel que $a^{2^i t} \equiv 1 \bmod n$, qui existe car $a^{2^s t} = a^{n-1} \equiv 1 \bmod n$ d'après le petit théorème de Fermat. On a $i \neq 0$ car a est un témoin. Donc l'entier $b = a^{2^{i-1} t}$ existe et vérifie $b^2 = a^{2^i t} \equiv 1 \bmod n$. Comme n est premier (et impair), les seules solutions de l'équation $x^2 = 1$ dans $\mathbb{Z}/n\mathbb{Z}$ sont 1 et -1. Or $b \not\equiv 1 \bmod n$ par minimalité de i, et $b \not\equiv -1 \bmod n$ car a est un témoin de n. C'est contradictoire.
4. ($\star$) Voir le Notebook ! 2 et 3 sont des témoins mais pas 50.
5. ($\star$) Voir le Notebook !
6. Par exponentiation rapide, on peut calculer x^t dans $\mathbb{Z}/n\mathbb{Z}$ en $O(\log t)$ multiplications dans $\mathbb{Z}/n\mathbb{Z}$, ce qui permet (en prenant pour x la classe de a) de tester si a^t est congru à 1 modulo n. Chaque multiplication dans $\mathbb{Z}/n\mathbb{Z}$ revient à effectuer une multiplication dans $\mathbb{Z}$, suivi d'une division euclidienne par n (puisqu'on représente chaque classe modulo n par son unique représentant compris entre 0 et $n-1$). Au total, le coût est donc de $O(\log t)$ opérations arithmétiques dans $\mathbb{Z}$.
7. Le calcul de a^t modulo n coûte $O(\log t)$ opérations arithmétiques dans $\mathbb{Z}$ d'après la question précédente. Ensuite on calcule les $a^{2^i t}$ de proche en proche, en réduisant modulo n à chaque étape, ce qui coûte $O(s)$ opérations arithmétiques dans $\mathbb{Z}$. Au total, le coût est donc de $O(s + \log t)$ opérations arithmétiques dans $\mathbb{Z}$. Or, quitte à supposer que les logarithmes sont en base 2, on a $s + \log t = \log(2^s t) \leq \log n$: le coût est donc de $O(\log n)$ opérations arithmétiques dans $\mathbb{Z}$.
8. La différence entre les deux versions est que dans la version 1, on calcule a^t dans $\mathbb{Z}$ alors que dans la version 2, on le calcule dans $\mathbb{Z}/n\mathbb{Z}$ (en réduisant modulo n après chaque multiplication). Bien évidemment la version 2 utilise davantage d'opérations, essentiellement le double, mais cette différence négligeable. Le vrai problème est que la version 1 oblige à manipuler des entiers extrêmement grands. Cela ralentit beaucoup le calcul : une opération arithmétique ne prend pas le même temps selon la taille des entiers mis en jeu, même si le plus souvent on ne tient pas compte de cela. Dans le contexte du problème, t et n sont en pratique du même ordre de grandeur, et extrêmement grands. Cela empêche en pratique de calculer a^t dans $\mathbb{Z}$: par exemple si n est de l'ordre de 2^{5000} (voir la fin de la partie 3), il n'y aucune chance de pouvoir stocker le résultat de ce calcul dans un ordinateur.

Partie 2 : Proportion de témoins

9. ($\star$) Voir le Notebook !
10. ($\star$) Dans cet exemple on voit que le théorème 1 est quasiment optimal : la proportion est proche de 3/4 (on obtient $\frac{9360}{12403} \approx 0,754656...$). Il y a cependant une subtilité : dans le théorème 1, il s'agit d'une proportion parmi les entiers a premiers avec n. Si on modifie la fonction `proportion` pour calculer la proportion de témoins parmi les entiers a premiers avec n, on trouve qu'il y en a exactement 3/4. Le théorème 1 est donc optimal. Voir le Notebook.
11. ($\star$) Voir le Notebook ! On trouve en moyenne une proportion de 98%, c'est beaucoup mieux que la minoration du théorème 1.

Partie 3 : Un test probabiliste

12. Si le programme renvoie "n est composé", on est certain que n est vraiment composé car on a trouvé un témoin de n (d'après la question 3). Si le programme renvoie "n est probablement premier", cela signifie simplement qu'il n'a pas trouvé de témoin après K essais. Si n est composé c'est peu probable si K est grand, puisque la proportion de témoins est assez grande : elle est $> 3/4$ grâce au théorème 1 (puisque les entiers a tels que $\text{pgcd}(a, n) \neq 1$ sont eux aussi des témoins de n, d'après la question 2).
13. ($\star$) Voir le Notebook !
14. ($\star$) On obtient que 12403 n'est pas premier. On a une chance sur 4 d'obtenir un faux positif ici, à savoir que l'algorithme renvoie `premier` alors que 12403 est composé. Avec $K = 10$, on obtient le même résultat mais la probabilité d'obtenir un faux positif est alors de $1/4^{10}$ et il faut en moyenne un million d'essais pour obtenir un faux positif.
15. Supposons réalisé l'évènement $[X \in \mathcal{P}_k]$. Cela correspond à un évènement élémentaire ω tel que $X(\omega)$ soit un nombre premier. Sous cette hypothèse, en appliquant le test à $X(\omega)$, on est certain que le test renverra "n est probablement premier" (voir question 12 : si le programme renvoie "n est composé", alors n est composé et ici ce n'est pas le cas). On a donc $P(T|X \in \mathcal{P}_k) = 1$.
16. On suppose maintenant que $X(\omega)$ est composé. Alors le test renvoie "n est probablement premier" si et seulement si, lors de chacun des K tirages d'entiers a, on a obtenu un entier a qui n'est pas un témoin. Lors de chaque tirage, la probabilité que cela arrive est majorée par 1/4 d'après le théorème 1 (et car les entiers a tels que $\text{pgcd}(a, n) \neq 1$ sont eux aussi des témoins de n). Les K tirages successifs pouvant être considérés comme indépendants (en supposant le générateur pseudo-aléatoire bien conçu !), on obtient $P(T|X \notin \mathcal{P}_k) \leq 4^{-K}$.
17. On a

$$P(X \in \mathcal{P}_k|T) = \frac{P(T \cap [X \in \mathcal{P}_k])}{P(T)} = \frac{P(T|X \in \mathcal{P}_k)P(X \in \mathcal{P}_k)}{P(T|X \in \mathcal{P}_k)P(X \in \mathcal{P}_k) + P(T|X \notin \mathcal{P}_k)P(X \notin \mathcal{P}_k)}$$

par la formule de Bayes (qu'on a ici redémontrée à partir de la définition d'une probabilité conditionnelle et de la formule des probabilités totales). D'après les questions 15 et 16 on a donc

$$P(X \in \mathcal{P}_k|T) = \frac{p_k}{p_k + (1 - p_k)P(T|X \notin \mathcal{P}_k)} \geq \frac{p_k}{p_k + (1 - p_k)4^{-K}}.$$

18. En pratique, on cherche à construire un nombre premier compris entre 2^k et 2^{k+1}. On tire au sort un nombre impair dans cet intervalle, selon la loi uniforme : c'est la variable aléatoire X. Si le test montre que cet entier est composé, on recommence jusqu'à obtenir un entier pour lequel le test affirme qu'il est probablement premier. On est alors dans la situation suivante : on sait que T est réalisé, et sachant cela on cherche la probabilité que X soit premier. Il s'agit bien de $P(X \in \mathcal{P}_k|T)$, dont on cherche à montrer qu'elle est très proche de 1.

19. On a
$$1 - P(X \in \mathcal{P}_k|T) \leq \frac{(1-p_k)4^{-K}}{p_k + (1-p_k)4^{-K}} \leq \frac{4^{-K}}{p_k}$$
ce qui donne $P(X \in \mathcal{P}_k|T) \geq 1 - 10^{-6}$ dès que $4^K p_k \geq 10^6$. En remplaçant p_k par sa valeur approchée $\frac{2}{k\ln(2)}$ (on peut en fait déduire du théorème des nombres premiers que $p_k \sim \frac{2}{k\ln(2)}$ quand $k \to +\infty$), on voit que prendre pour K la partie entière supérieure de
$$\frac{\ln(k \cdot 10^6 \ln(2)/2)}{\ln 4} = \frac{\ln k}{\ln 4} + \frac{6\ln(10) + \ln(\ln(2)) - \ln(2)}{\ln 4} \simeq 0.721\ln(k) + 9.201 \qquad (1)$$
permet d'assurer que $P(X \in \mathcal{P}_k|T) \geq 1 - 10^{-6}$. Cette valeur est relativement faible, et elle augmente très lentement avec k. D'après la question 7, le coût du test est de $O(K\log(2^k)) = O(K \cdot k)$ opérations arithmétiques puisqu'on peut choisir aléatoirement un entier $a \in [\![2, n-1]\!]$ avec le même coût que $O(1)$ opérations arithmétiques. C'est tout à fait raisonnable !

20. Une clef RSA est le produit de deux grands nombres premiers, qu'on peut choisir de l'ordre de 2^{5000} par exemple. Avec les notations de la question précédente, on peut donc prendre $k = 5000$ et la formule (1) donne alors $K = 16$. D'après la question précédente, le coût du test de primalité est d'environ $500 \cdot 16 = 8000$ opérations arithmétiques. Bien sûr il s'agit juste d'un ordre de grandeur (on n'est pas à un facteur 5 ou 10 près...)

21. Pour générer un nombre premier de l'ordre de 2^{5000}, on simule X, on teste sa primalité, et on recommence jusqu'à ce que le test dise que X est probablement premier. Le nombre d'essais suit alors une loi géométrique de paramètre $p_k \simeq \frac{2}{k\ln(2)} \simeq \frac{1}{1733}$. En moyenne, il faut donc 1733 essais pour trouver ainsi un nombre premier ; compte tenu de la question précédente, cela coûte $1733 \cdot 8000 \simeq 14 \cdot 10^6$ opérations arithmétiques. Finalement, le coût moyen pour générer une clef RSA est de quelques dizaines de millions d'opérations arithmétiques : ça se fait très vite. Le risque d'erreur est majoré par 10^{-6} avec les choix faits ici, et on peut le réduire en augmentant K.

Partie 4 : Puissances dans un groupe cyclique

22. S'il existe $g \in G$ tel que $g^m = g_0$, alors on a $g_0^{\varphi/k} = g^{m\varphi/k} = (g^\varphi)^{m/k} = 1$ car $g^\varphi = 1$ et $k|m$. Réciproquement, quitte à fixer un isomorphisme entre G et $\mathbb{Z}/\varphi\mathbb{Z}$ on peut supposer que $G = \mathbb{Z}/\varphi\mathbb{Z}$ (noté additivement). On suppose $(\varphi/k)g_0 = 0$ c'est-à-dire que $g_0 = \overline{kg_1}$ pour un certain $g_1 \in \mathbb{Z}$, en notant $\overline{a}$ la classe modulo φ d'un entier a. En notant $m = km'$ et $j = kj'$, l'équation $mg = g_0$ s'écrit $km'g = \overline{kg_1}$. Comme m' et φ/k sont premiers entre eux, il existe $u, v \in \mathbb{Z}$ tels que $m'u = 1 + v\varphi/k$, d'où $km'ug_1 = kg_1 + v\varphi g_1 \equiv kg_1 \bmod \varphi$.

Donc $g = \overline{ug_1}$ est une solution. Les autres sont les $g = \overline{ug_1 + h}$ avec $h \in \mathbb{Z}$ tel que $km'h \equiv 0 \bmod \varphi$, ce qui équivaut à $m'h \equiv 0 \bmod \varphi/k$ et à $h \equiv 0 \bmod \varphi/k$ puisque m' et φ/k sont premiers entre eux. Il y a donc exactement k valeurs de h (modulo φ) qui conviennent.
On donne ici une méthode alternative (qui revient en fait au même). On note a un générateur de G et on écrit $g_0 = a^{n_0}$ et on cherche les solutions g sous la forme $g = a^n$. Il s'agit de trouver n modulo φ tel que $a^{mn} = a^{n_0}$ soit tel que $mn \equiv n_0 \,(\mathrm{mod}\ \varphi)$. Si $g_0^{\varphi/k} = 1$, on a alors l'existence de n_1 tel que $n_0 = kn_1$ et en écrivant $m = km'$ et $\varphi = k\varphi'$, il vient $m'n \equiv n_1 \,(\mathrm{mod}\ \varphi')$. Comme m' est premier avec φ', cela détermine complétement n modulo φ' et fournit $\varphi/\varphi' = k$ choix modulo φ.

23. Le résultat découle immédiatement de la question 22 appliquée à $g_0 = 1$, $m = t$ et $k = w$, qui vérifient bien $g_0^{\varphi/k} = 1$.

24. On applique à nouveau la question 22, cette fois avec $g_0 = \alpha$ et $m = 2^j t$. Si $0 \leq j \leq u-1$ alors $k = 2^j w$ et $\varphi/k = 2^{u-j}t/w$ est pair, donc on a $g_0^{\varphi/k} = 1$ et l'équation $g^{2^j t} = \alpha$ possède exactement $k = 2^j w$ solutions. En revanche, si $j \geq u$ alors $k = 2^u w$ et $\varphi/k = t/w$ est impair, donc $g_0^{\varphi/k} = \alpha \neq 1$ et l'équation $g^{2^j t} = \alpha$ n'a aucune solution.

Partie 5 : Preuve du théorème 1

25. Pour tout i, l'équation $x^t = 1$ possède w_i solutions dans $(\mathbb{Z}/p_i^{a_i}\mathbb{Z})^\times$ d'après la question 23. D'après le théorème chinois, elle en possède donc $w_1 \cdot \ldots \cdot w_N = V/V'$ dans $(\mathbb{Z}/n\mathbb{Z})^\times$.

26. On remarque déjà que $\alpha = -1$ est d'ordre 2 dans $(\mathbb{Z}/p_i^{a_i}\mathbb{Z})^\times$ pour tout i, car les p_i sont impairs. Si $j \geq \min(u_1, \ldots, u_N)$, il existe i tel que $j \geq u_i$. D'après la question 24, l'équation $x^{2^j t} = -1$ ne possède aucune solution dans $(\mathbb{Z}/p_i^{a_i}\mathbb{Z})^\times$: c'est donc aussi le cas dans $(\mathbb{Z}/n\mathbb{Z})^\times$. En revanche, si $j \leq \min(u_1, \ldots, u_N)$ alors pour tout i, la question 24 montre que l'équation $x^{2^j t} = -1$ possède $2^j w_i$ solutions dans $(\mathbb{Z}/p_i^{a_i}\mathbb{Z})^\times$; d'après le théorème chinois, elle en possède donc $2^{jN} w_1 \cdot \ldots \cdot w_N = 2^{jN} V/V'$ dans $(\mathbb{Z}/n\mathbb{Z})^\times$.

27. Pour $0 \leq j \leq s-1$ notons E_j l'ensemble des $x \in (\mathbb{Z}/n\mathbb{Z})^\times$ tels que $x^{2^j t} = -1$, et aussi E_{-1} l'ensemble des $x \in (\mathbb{Z}/n\mathbb{Z})^\times$ tels que $x^t = 1$. Ces ensembles sont deux à deux disjoints, et leur réunion est l'ensemble E des classes des non-témoins. On a $\operatorname{Card} E_{-1} = V/V'$ d'après la question 25, $\operatorname{Card} E_j = 2^{jN} V/V'$ pour $0 \leq j \leq r-1$ et $\operatorname{Card} E_j = 0$ pour $r \leq j \leq s-1$ d'après la question 26, d'où

$$\operatorname{Card} E = \Big(1 + \sum_{j=0}^{r-1} (2^N)^j\Big) \frac{V}{V'} = A.$$

28. Lorsque $N = 1$ on a $A = 2^r V/V' = 2^r w_1$ avec $p_1^{a_1} - 1 = 2^s t$ et $\varphi_1 = p_1^{a_1-1}(p_1 - 1) = 2^{u_1} v_1$. Comme $p_1^{a_1-1}$ est impair et que $p_1 - 1$ divise $p_1^{a_1} - 1$, on a $u_1 \leq s$ et $r = u_1$. La proportion de non-témoins parmi les entiers $a \in [\![2, n-1]\!]$ premiers avec n est donc $A/\varphi = w_1/v_1 = \operatorname{pgcd}(t, v_1)/v_1$. Comme n est composé, on a $a_1 \geq 2$ et $p_1^{a_1-1}$ divise v_1, alors que $p_1^{a_1-1}$ est premier avec $p_1^{a_1} - 1$ donc avec t. On a donc $\operatorname{pgcd}(t, v_1) \leq v_1/p_1^{a_1-1}$, c'est-à-dire $A/\varphi \leq 1/p_1^{a_1-1}$. Comme p_1 est impair et $a_1 \geq 2$, cela implique $A/\varphi \leq 1/4$ sauf dans un cas : celui où $p_1 = 3$ et $a_1 = 2$. Mais dans ce cas, on a $n = 9$ alors qu'on suppose $n \geq 11$. Le théorème 1 est donc démontré si $N = 1$.

Supposons maintenant $N \geq 2$. Alors on a $2^N - 1 > 2^{N-1}$ donc $AV'/V < 1 + 2^{Nr-N+1}$. Or AV'/V est entier d'après la preuve de la question précédente, donc $AV'/V \leq 2^{Nr-N+1}$ ce qui donne

$$A/\varphi(n) \leq 2^{Nr-U} 2^{1-N}/V' \leq \frac{1}{2^{N-1}V'}$$

puisque $U \geq Nr$. Supposons maintenant, par l'absurde, que $A/\varphi(n) > 1/4$. Alors on a $2^{N-1}V' < 4$ d'où $N = 2$ et $V' = 1$, ce qui donne $v_1' = v_2' = 1$: les entiers v_1 et v_2 divisent t.

Soit $i \in \{1, 2\}$ tel que $a_i \geq 2$. Alors $p_i^{a_i - 1}$ divise v_i mais pas w_i (car sinon p_i diviserait t, donc $n-1$, et aussi n). Donc $p_i^{a_i-1}$ divise $v_i' = 1$: c'est impossible. On a donc $a_1 = a_2 = 1$. En reprenant la majoration $A/\varphi(n) \leq 2^{Nr-U} 2^{1-N}/V'$ on obtient $1/2 < 2^{2r-U}$ d'où $U = 2r$ ce qui impose $u_1 = u_2 = r$. Finalement on a $\varphi_i = p_i - 1 = 2^r v_i$ avec $v_i | t$ pour tout $i \in \{1, 2\}$. Or $r \leq s$ donc $p_i - 1$ divise $n - 1 = (p_1 - 1)(p_2 - 1) + (p_1 - 1) + (p_2 - 1)$. On en déduit que $p_1 - 1$ et $p_2 - 1$ se divisent mutuellement : c'est impossible.

29. ($\star$) On a $n = 12403 = p_1 p_2$ avec $p_1 = 79$ et $p_2 = 157$ premiers, et une propriété particulière : $p_2 = 2p_1 - 1$. On a $N = 2$, $\varphi_1 = p_1 - 1 = 78$ et $\varphi_2 = p_2 - 1 = 156$ donc $u_1 = 1$, $u_2 = 2$, $v_1 = v_2 = 39$, et on a $r = s = 1$, $t = 6201 = 3^2 \cdot 13 \cdot 53$ d'où $w_1 = w_2 = 39$, $V = 39^2$ et $V' = 0$. On a donc $A = 2 \cdot 39^2$ et $A/\varphi = 1/4$. Cela confirme ce qu'on a constaté à la question 10 : il s'agit d'un cas d'égalité dans le théorème 1. La situation est la même pour tout entier n de la forme $p_1(2p_1 - 1)$, où $p_1 \equiv 3 \bmod 4$ est un nombre premier tel que $2p_1 - 1$ soit premier. En effet on a toujours $N = 2$, $\varphi_1 = p_1 - 1 \equiv 2 \bmod 4$ et $\varphi_2 = p_2 - 1 = 2(p_1 - 1)$ donc $u_1 = 1$, $u_2 = 2$, $v_1 = v_2 = (p_1 - 1)/2$. Comme $N = \equiv 3 \bmod 4$ on a $r = s = 1$, $t = (n-1)/2 = (2p_1^2 - p_1 - 1)/2 = (2p_1 + 1) \cdot (p_1 - 1)/2$ d'où $w_1 = w_2 = (p_1 - 1)/2$, $V = (p_1 - 1)^2/4$ et $V' = 0$. Finalement on a donc $A = 2(p_1 - 1)^2/4$ et $A/\varphi = 1/4$. D'autres exemples sont 15, 91, 703, 1891, 38503,... Voir le Notebook !