De la rareté au tsunami de données

Une histoire de la séparation aveugle de sources

Matthieu Puigt
Université Littoral Côte d'Opale, LISIC – UR4491

Nous revisitons l'histoire de la séparation de sources au regard des volumes de données disponibles : nous montrons comment ces derniers ont permis le développement de techniques d'apprentissage (profond). En conclusion, nous présentons quelques perspectives prenant en compte les contraintes environnementales, énergétiques et matérielles qui surviendront dans un futur plus ou moins proche.

Supposons que votre petit voisin revient du cinéma où il a vu un film d'animation sur un groupe de chameaux et de dromadaires. Il vous explique qu'il a compté 5 têtes et qu'ils avaient au total 7 bosses. Si on souhaite retrouver le nombre de dromadaires et de chameaux à partir de ces informations, on peut faire appel à plusieurs stratégies. Ici, compte tenu des nombres en jeu, on peut facilement procéder par essai-erreur. Sinon — si on se rappelle de ses cours de mathématiques de lycée (ou de collège pour les moins jeunes) — on peut aussi poser un système d'équations en se rappelant qu'un chameau a deux bosses, un dromadaire une seule (et que chaque animal n'a qu'une seule tête) soit, en notant c le nombre de chameaux et d le nombre de dromadaires :

$$c + d = 5 \qquad \text{et} \qquad 2c + d = 7 \tag{1}$$

Imaginons maintenant que le film que votre petit voisin a vu se passe dans un monde imaginaire. Aussi, même si vous savez toujours qu'il y a 2 espèces différentes, 5 têtes et 7 bosses, vous ne savez combien de têtes ni de bosses possède chaque espèce… Dans ce cas, les coefficients devant c et d dans l'équation (1) sont eux-aussi inconnus. Le problème ainsi énoncé n'est pas soluble. C'est pourtant ce que cherche à résoudre la *séparation aveugle de sources* (SAS).

DOI:10.48556/SIF.1024.25.149

Cette discipline, née au début des années 80 en France [28], cherche à estimer un ensemble de p sources inconnues notées[1] $s_j(t)$ à partir de n mélanges $x_i(t)$ de ces signaux, les paramètres de mélange des signaux sources jusqu'aux capteurs étant eux-aussi inconnus[2]. En regroupant les sources et les observations à l'instant t sous formes de vecteurs colonnes, notés respectivement $\mathbf{s}(t)$ et $\mathbf{x}(t)$, et en notant $\mathcal{A}$ l'opérateur de mélange inconnu, nous obtenons la relation[3] :

$$\mathbf{x}(t) \approx \mathcal{A}\left(\mathbf{s}(t)\right). \tag{2}$$

Plusieurs modèles — linéaires ou (post-)non-linéaires [46], [47] — ont été proposés pour exprimer l'opérateur $\mathcal{A}$ [17]. Plus particulièrement, pour le type de mélange le plus simple considéré dans la littérature, c.-à-d. les mélanges linéaires instantanés, le problème de SAS s'écrit

$$x_i(t) \approx \sum_{j=1}^{p} a_{ij} s_j(t), \tag{3}$$

où le paramètre a_{ij} est le coefficient de contribution de la source j dans l'observation i. En supposant que les signaux soient discrétisés et que nous ayons m échantillons de chaque observation, nous pouvons représenter les données observées sous la forme d'une matrice X de taille $n \times m$. De même, on peut représenter les signaux sources à estimer sous la forme d'une matrice S de taille $p \times m$ et les coefficients de mélange sous la forme d'une matrice A de taille $n \times p$. L'équation (3) s'écrit alors sous la forme :

$$X \approx A \cdot S. \tag{4}$$

Pour un modèle de mélange convolutif qui étend (3) (en remplaçant le paramètre a_{ij} par un filtre $a_{ij}(t)$ et le produit par une convolution), il est encore possible d'écrire le problème sous la forme (4) soit dans le domaine temporel (dans ce cas, la matrice A est Toeplitz par bloc [12]) soit dans le domaine fréquentiel (dans ce cas, pour chaque pulsation considérée, le problème de SAS peut s'écrire sous la forme (4), voir [41]). Enfin, dans le cas où des non-linéarités apparaissent dans les mélanges, une stratégie qu'on retrouve en apprentissage de variété [54] et dans certaines méthodes de SAS [47] consiste à supposer que ces non-linéarités peuvent être *localement* considérées comme

1. On adopte ici les notations lorsque le problème de SAS est lié à un problème audio. Les variables dépendent alors du temps, d'où le choix de la variable t. Cependant, le problème de SAS est générique et peut notamment s'exprimer avec des variables multi-dimensionnelles.
2. Cependant, on suppose connaître le modèle de mélange (mais pas les paramètres de ce modèle). De plus, même si certains auteurs proposent de l'estimer [4], le nombre p de sources est généralement connu lui-aussi.
3. L'égalité dans l'équation (2) n'est pas stricte en raison de la présence de bruit, que nous ne modélisons pas dans cet article.

Une hypothèse clé de la SAS réside dans la *diversité* de l'information. En effet, en reprenant l'exemple du petit voisin en début d'article, s'il fournit le nombre de têtes et de pattes du groupe[4], les deux équations du système seraient identiques à un facteur 4 prés, puisque chaque animal possède 4 pattes. En pratique, la diversité en SAS peut être par exemple spatiale (des capteurs positionnés en des lieux différents collectent des mélanges avec des contributions différentes des mêmes sources) ou être propre aux sources (par exemple, des représentations parcimonieuses différentes [23]).

La SAS a intéressé un nombre important de chercheurs issus de différentes communautés (traitement statistique du signal et des images, apprentissage, etc.) et un nombre important d'applications comme l'audio, le bio-médical, l'imagerie (hyperspectrale), les données bancaires ou l'analyse chimique [17, chap. 16]. Dans cet article, nous ne rappelons pas seulement l'histoire de cette discipline au cours du temps (nous renvoyons le lecteur à [28] pour découvrir les travaux pionniers) mais nous la positionnons à l'aune des quantités de données disponibles. En effet, aujourd'hui , une partie de la communauté a pris le virage de l'apprentissage profond, tirant parti de volumes de données toujours plus grands pour entraîner des modèles complexes.

Cet article est structuré ainsi. La section « Au commencement : des données éparses » introduit les approches historiques de SAS. Nous soulignons les liens entre SAS et d'autres domaines, sous un chapeau commun d'analyse en variables latentes, dans la section « De la SAS à l'analyse en variables latentes » puis le virage de l'apprentissage et l'apprentissage profond dans la section « Le virage de l'apprentissage (profond) ». Enfin, nous concluons et discutons de l'avenir de la discipline dans la section « Conclusion et un peu de futurologie *Business as usual* ou prise en compte de limites ? »… en tenant compte des contraintes environnementales et des contraintes de ressources.

Au commencement : des données éparses

Les débuts de la SAS coïncident avec le développement de méthodes d'analyse en composantes indépendantes (ICA, pour *Independent Component Analysis*). Celles-ci combinent les données observées $\mathbf{x}(t)$ via un opérateur $\mathcal{B}$ de sorte que les signaux en sortie de $\mathcal{B}$, notés

$$\mathbf{y}(t) = \mathcal{B}(\mathbf{x}(t)) \tag{5}$$

soient statistiquement indépendants. Si les signaux sources sont eux-même statistiquement indépendants, alors ils sont égaux aux signaux de sortie, à certaines ambiguïtés prés. Dans le cas d'un modèle de mélange linéaire instantané (4), les opérateurs $\mathcal{A}$ et $\mathcal{B}$ sont de simples matrices notées respectiv-

4. On suppose qu'aucune espèce ne possède d'infirmité discriminante.

ement A et B et cette dernière est égale à la (pseudo-)inverse de A, à une permutation et un facteur d'échelle près. Les signaux de sortie s'écrivent alors

$$\mathbf{y}(t) = B \cdot \mathbf{x}(t) \approx B \cdot A \cdot \mathbf{s}(t) \approx \Lambda \cdot P \cdot \mathbf{s}(t) \tag{6}$$

où Λ est une matrice diagonale et P est une matrice de permutation.

Les travaux pionniers en ICA ont porté sur l'identifiabilité de l'opérateur $\mathcal{B}$, pour des mélanges linéaires [16] et non-linéaires [26], ainsi que sur le développement de méthodes pour réaliser cette tâche. Parmi les nombreux travaux marquants, P. Comon a démontré dans [16] qu'il était possible de séparer des signaux statistiquement indépendants à partir de signaux mélangés, sous réserve qu'au plus une de ces sources soit gaussienne. Ces approches sont non-supervisées et n'utilisent que les données observées pour estimer l'opérateur $\mathcal{B}$. A la fin des années 90, il est cependant compliqué d'avoir des jeux de données réels et, afin d'organiser des comparaisons objectives de méthodes, la communauté cherche à regrouper codes et jeux de données sur un unique site web ICA Central et à échanger sur une liste de diffusion[5].

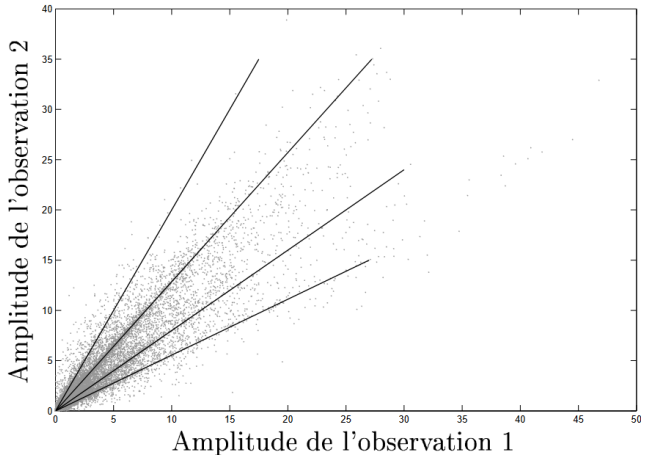

Fig. 1. Nuage de points et paramètres de mélange (droites) pour un mélange stéréo (n = 2) de p = 4 sources de parole. Chaque point en gris correspond à la valeur du module de la transformée TF d'une observation en un point TF donné par rapport à celui de l'autre observation (source [44]).

C'est à la même période que des hypothèses alternatives à l'indépendance statistique ont été étudiées pour réaliser la SAS. D'une part, des méthodes fondées sur la parcimonie ont été proposées. Elles s'appuient initialement sur

5. Tous deux ont aujourd'hui disparu et ont été remplacés par un autre site (`https://lva-central.irisa.fr/`) et un groupe google (`https://groups.google.com/g/lvalist`).

les propriétés de localité des signaux de parole et de musique dans le domaine temps-fréquence (TF). Plus particulièrement, deux visions coexistent. D'un côté, certains auteurs s'intéressent aux décompositions TF énergétiques et proposent des approches de SAS qui s'inspirent des méthodes statistiques d'ordre 2, voir [7]. Au contraire, une vaste majorité de chercheurs s'intéresse au décompositions atomiques, comme la transformée de Fourier à court terme par exemple [23]. Celles-ci présentent l'avantage d'être linéaires, transformant un problème linéaire de SAS où les sources sont fortement recouvrantes entre elles en un nouveau problème où les sources sont parcimonieuses et potentiellement disjointes dans le domaine transformé [62] (voir figure 1).

Compte tenu de la parcimonie des sources, les observations sont alors localement [62], [2] de faible rang et il devient possible de résoudre les problèmes sous-déterminés, c.-à-d. où le nombre p de sources est supérieur au nombre n d'observations. Dans ce cas, il est nécessaire que le nombre de sources actives soit (strictement) inférieur à n dans chaque point TF[6]. Ces approches fonctionnent en deux temps : les paramètres de mélange sont tout d'abord estimés puis les sources sont reconstruites. Cette dernière étape est obtenue par masquage temps-fréquence [62] ou en résolvant des problèmes localement inversibles [56], mais cela se fait au prix de bruit musical perceptible pour l'oreille humaine.

D'autre part, des approches fondées sur la positivité des signaux observés et à estimer ont été proposées [32]. Ces méthodes, nommées NMF (pour *Nonnegative Matrix Factorization*) consistent à résoudre l'équation (4) sous l'hypothèse que les matrices X, A et S sont à valeurs positives ou nulles. Appliquée à un enregistrement audio, la NMF ne sépare pas directement les sources mais cherche une approximation de faible rang d'un spectrogramme (c.-à-d. le module de la transformée TF d'un enregistrement audio) de sorte qu'une matrice facteur contienne les signatures fréquentielles de sons et l'autre les différents coefficients d'activation temporelle de ces signatures (voir figure 2). La phase est ensuite estimée par filtrage des signaux originaux. Dans le cas de la SAS avec plusieurs observations, une étape intermédiaire consiste à estimer quelles signatures fréquentielles et coefficients d'activations temporelles sont associés à quelles sources [40]. De nombreuses extensions de la NMF ont été proposées, notamment les décompositions tensorielles non-négatives qui étendent la NMF à des données de plus grande dimension [63].

6. Bien que proposées de manière indépendantes, ces méthodes partagent de nombreuses propriétés des techniques de classification de sous-espaces [55].

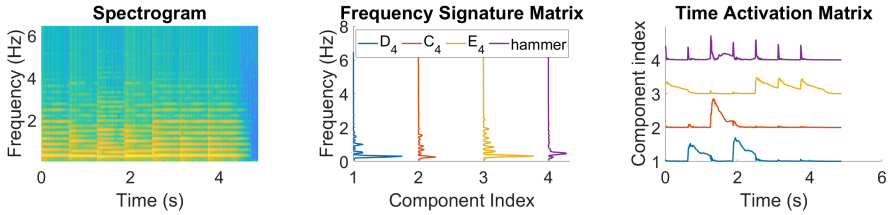

Fig. 2. Exemple de décomposition NMF sur un morceau joué au piano. La NMF permet de décomposer le contenu fréquentiel en 3 notes et la signature des marteaux (adapté de Nicolas Gillis [22]).

La NMF connaît depuis 25 ans un grand intérêt de la part des communautés signal, image et apprentissage, bien au-delà de la SAS. Son développement coïncide avec celui d'autres problèmes au formalisme proche.

De la SAS à l'analyse en variables latentes

L'expression (4) que nous avons énoncée en introduction n'est pas propre à la SAS mais au contraire commune à de nombreux problèmes de traitement de données (signal, image) et d'apprentissage statistique. En fonction de la mesure $\mathcal{D}(X, A \cdot S)$ de distance[7] entre X et le produit $A \cdot S$ et en fonction de contraintes sur A ou S, l'équation (4) permet de modéliser de nombreux problèmes, comme on peut le voir sur le tableau 1. Tous ont en commun l'estimation de variables latentes ou cachées dans les données, avec des propriétés diverses. Par ailleurs, ces méthodes ne sont pas nécessairement appliquées à la SAS, d'où le nom d'analyse en variables latentes (LVA, pour *Latent Variable Analysis*) que l'on retrouve de plus en plus souvent dans la communauté. Enfin, ce type de problème intéresse à la fois les communautés en traitement du signal et des images et en apprentissage statistique.

Pour illustrer ce phénomène, il est intéressant de remarquer l'évolution de la conférence de SAS organisée pour la première fois à Aussois, en France, en 1999. Elle s'appelait alors ICA puisque historiquement, seules les méthodes d'ICA permettaient de traiter le problème de SAS (pour rappel, les premières méthodes de NMF et de SCA naissent entre 1999 et 2001). En 2010, la conférence devint LVA-ICA pour montrer l'ouverture de la communauté vers l'analyse en variables latentes. En 2013 et 2014, la conférence fut remplacée par une session spéciale de la conférence IEEE MLSP (*Machine Learning for Signal Processing*), montrant le virage de la communauté vers l'apprentissage.

7. Les distances mises en avant dans le tableau 1 sont la norme de Frobenius, la norme ℓ_1 et la divergence de Kullback-Leibler respectivement notées $\|\cdot\|_{\mathcal{F}}$, $\|\cdot\|_1$ et $KL(\cdot,\cdot)$.

Méthode	$\mathcal{D}(X, A \cdot S)$	Contrainte sur A	Contrainte sur S
SVD	$\|X - A \cdot S\|_{\mathcal{F}}^2$	$A^T A = I$	$S^T S = \Lambda$
k-means	$\|X - A \cdot S\|_{\mathcal{F}}^2$	sans	$S^T S = I$ $s_{ij} = \{0, 1\}$
k-medians	$\|X - A \cdot S\|_1$	sans	$S^T S = I$ $s_{ij} = \{0, 1\}$
pLSI	$KL(X, A \cdot S)$	$\mathbb{1}^T \cdot A \cdot \mathbb{1} = 1$ $a_{ij} \geq 0$	$\mathbb{1}^T \cdot S = \mathbb{1}$ $s_{ij} \geq 0$
NMF	$\|X - A \cdot S\|_{\mathcal{F}}^2$ ou $KL(X, A \cdot S)$	$a_{ij} \geq 0$	$s_{ij} \geq 0$
MMV	$\|X - A \cdot S\|_{\mathcal{F}}^2$	sans	parcimonie structurée
CX	$\|X - A \cdot S\|_{\mathcal{F}}^2$	colonnes de A extraites de X	sans
NMF convexe	$\|X - A \cdot S\|_{\mathcal{F}}^2$	colonnes de A extraites de X	$s_{ij} \geq 0$
NMF convexe robuste	$\|X - A \cdot S\|_1$	colonnes de A extraites de X	$s_{ij} \geq 0$

Table 1. Différents problèmes de factorisation matricielle (partiellement adapté de [6], [14]).

Un des problèmes les plus célèbres d'analyse en variables latentes concerne les systèmes de recommandation. Il s'agit en pratique d'inférer les données d'une matrice partiellement observée. Popularisées par le concours organisé par Netflix à la fin des années 2000, les approches fondées sur la factorisation matricielle se sont montrées extrêmement bien adaptées pour gérer ce type de problème [30]. En notant X et W les matrices de données et de poids (w_{ij} valant 1 si x_{ij} est connu et 0 sinon) et $\circ$ l'opérateur de Hadamard[8], cela revient à résoudre

$$W \circ X \approx W \circ (A \cdot S). \tag{7}$$

Une fois A et S estimées, les données manquantes de X peuvent être prédites par le produit $A \cdot S$. Le problème (7) a aussi été rencontré pour d'autres applications, comme par exemple l'étalonnage *in situ* de capteurs mobiles [19] ou le dématriçage et le démélange conjoint d'images multispectrales [1].

8. Produit de matrices de mêmes dimensions, coefficient à coefficient.

Le virage de l'apprentissage (profond)

Dès les années 1990-2000, les statisticiens et les spécialistes de traitement du signal et des images s'intéressent aux approximations parcimonieuses. Elles consistent à expliquer des données à partir d'un nombre réduit d'*atomes* contenus dans un *dictionnaire*. Les années 90 voient l'essor de dictionnaires fixes, tels que les transformées en ondelette. Mais rapidement, l'idée d'apprendre des atomes à partir de jeux de données émerge. De nombreuses applications voient le jour, comme la compression ou la restauration de signaux (déconvolution, débruitage) [20]. L'apprentissage de dictionnaire peut être modélisé sous la forme (4) — avec une matrice X potentiellement de très grande taille — et l'hypothèse de parcimonie est traitée sous la forme d'une pénalisation de la norme ℓ_1 d'une matrice [36] : à partir d'une collection de données de références organisées sous forme d'une matrice X, on apprend les atomes dans la matrice[9] A et les coefficients d'activation dans une matrice creuse S. C'est à cette période que la quantité de données disponibles pour l'apprentissage explose, certains auteurs parlant même de déluge de données [5]. De nombreuses stratégies sont alors proposées : calculs distribués [34], en ligne [36] ou comprimés [60].

Dans le domaine audio, la NMF va grandement bénéficier de cette stratégie d'apprentissage supervisé : en appliquant la NMF sous contrainte de parcimonie sur des spectrogrammes de signaux de référence (c.-à-d. des signaux sources non-mélangés), il est possible d'apprendre des signatures fréquentielles qui seront figées dans un des facteurs matriciels de la NMF dédiée à la SAS [50]. Au début des années 2010, lorsque les approches d'apprentissage profond émergent [13], une partie de la communauté SAS bascule vers ce formalisme, en remplaçant la NMF par des réseaux de neurones profonds [38]. Cette communauté se regroupe en créant sa propre liste[10] Machine Listening en 2012. Aujourd'hui, de très nombreuses méthodes utilisant différentes architectures de réseaux ont été proposées [48], [59].

Pour accompagner ce changement de paradigme, il est intéressant de voir l'évolution des jeux de données durant cette période. En effet, dès le milieu des années 2000, des concours ou des campagnes d'évaluation de méthodes de SAS sont organisées : *Speech Separation Challenge* du réseau PASCAL en 2006[11], concours de méthodes de SAS pour des mélanges convolutifs lors de la conférence MLSP en 2007[12] et surtout les campagnes d'évaluation lors des conférences ICA/ LVA-ICA (SASSEC en 2007, SiSEC entre 2008 et 2018) [57].

9. En transposant X, les rôles des matrices A et S peuvent être permutés.
10. Voir `https://groups.google.com/g/machinelistening`.
11. Voir l'archive web : `https://web.archive.org/web/20090101151757/http://www.dcs.shef.ac.uk/~martin/SpeechSeparationChallenge.htm`.
12. Voir l'archive web : `https://web.archive.org/web/20070530031451/https://mlsp2007.conwiz.dk/index.php?id=43`.

Lorsqu'on s'intéresse au volume de signaux musicaux enregistrés en studio, la communauté proposait seulement 2 jeux de données pour l'entraînement (et 2 pour le test) en 2008[13], contre 100 jeux pour l'entraînement (et 50 pour le test) dix ans plus tard[14].

A l'image des challenges CHiME[15] — qui ont pris la suite du challenge de SAS du réseau PASCAL tout en élargissant son périmètre au traitement de la parole dans des environnements divers (avec, pour chaque tâche de chaque challenge, plusieurs heures de données d'entraînement) — une tendance actuelle consiste à développer des outils d'intelligence artificielle qui réalisent plusieurs tâches, par exemple la segmentation, la séparation et la transcription automatique de la parole, à l'aide par exemple de modèles de langage de grande dimension [24]. Une autre tendance cherche à à fournir une alternative explicable aux réseaux profonds qui fonctionnent comme des boîtes noires logicielles. Parmi ces alternatives, l'*unrolling* s'appuie sur des outils classiques en traitement du signal et des images tels que la NMF [37].

Conclusion et un peu de futurologie *Business as usual* ou prise en compte de limites ?

Dans cet article, nous avons revisité l'histoire de la SAS au regard des volumes de données disponibles. Pour illustrer ce point de vue, nous avons choisi le domaine audio comme champ applicatif mais notre analyse reste vraie dans d'autres domaines où la SAS est très étudiée, par exemple le démélange hyperspectral en télédétection (où les approches historiques [10] ont en partie laissé la place aux méthodes d'apprentissage profond [8]). Nous proposons maintenant d'esquisser quelques perspectives à moyen terme, compte tenu des contraintes de ressources (matières première, énergie) qui vont impacter l'humanité.

Malgré son intérêt pour, par exemple, l'optimisation de processus industriels ou agricoles ou le développement de nouveaux médicaments, l'IA nécessite des besoins toujours plus importants en données et en capacité de calcul, et donc en énergie. Ainsi, en prenant l'exemple des modèles génératifs, le seul entraînement du modèle GPT-3 nécessiterait plus de 400 ans de calculs sur un ordinateur de bureau et produirait 552 tonnes équivalent-carbone de CO_2 [42]. Les auteurs du rapport [49] prédisaient dès 2015 que la consommation énergétique mondiale du secteur informatique pourrait dépasser la capacité mondiale de production d'énergie disponible à l'horizon 2040–2050… Des études plus récentes [21] confirment cette tendance. Même si des efforts réels ont été

13. http://sisec2008.wiki.irisa.fr/tiki-index165d.html?page=Professionally+produced+music+recordings.
14. Voir : https://sisec.inria.fr/2018-professionally-produced-music-recordings/.
15. Voir https://www.chimechallenge.org/.

réalisés — par exemple pour optimiser les *data centers* — l'effet rebond, c.-à-d. l'accroissement des usages rendu possible par l'amélioration de l'efficacité énergétique, en annule tout bénéfice [27]. De plus, de nombreux défis environnementaux attendent le monde, notamment le réchauffement climatique et la transition énergétique (ne serait-ce que pour remplacer les énergies fossiles par des énergies bas carbone). Cette transition passe aujourd'hui par un usage croissant de métaux, dont les concentrations dans les mines diminuent significativement depuis plusieurs décennies [51]. Dans un monde aux limites finies, une croissance infinie des usages de ressources n'est pas possible.

On peut dès lors se demander s'il est vraiment opportun de continuer à développer tous azimuts des méthodes toujours plus complexes pour un nombre toujours plus important d'applications. Au contraire, peut-être devrions-nous collectivement faire preuve de techno-discernement [9], c.-à-d. corréler l'usage de solutions informatiques énergivores à la valeur ajoutée sociétale ou environnementale qu'elles produisent. La définition du périmètre de cette valeur ajoutée est loin d'être simple, et même si un consensus se dégageait autour de certaines thématiques, nous devrions probablement arrêter de faire mieux quoi qu'il en coûte mais chercher à faire (presque) aussi bien qu'aujourd'hui avec un usage de plus en plus réduit de ressources. Cette approche nécessite de prendre en compte toutes les externalités négatives liées aux usages numériques (par exemple, l'usage de l'eau [33] ou le coût énergétique [31]) et probablement changer certains paradigmes. Nous soulignons quelques pistes ci-dessous.

Changer le hardware

Le silicium, c.-à-d. le matériau de base des transistors et des circuits intégrés, a non-seulement permis la prospérité de la Silicon Valley mais aussi le développement de l'informatique moderne. Cependant, de nombreuses alternatives à l'ordinateur personnel tels que nous le concevons aujourd'hui existent : sans remonter jusqu'aux calculateurs mécaniques de Pascal et sans parler d'informatique quantique, les calculateurs analogiques [11], [53] ont permis d'effectuer des calculs scientifiques durant une grande partie du xx^{e} siècle. Compte tenu de la fin de la loi de Moore (la vitesse de calcul des processeurs a arrêté de doubler tous les deux ans) [52] et de leur frugalité énergétique, on s'intéresse à nouveau à ces hardwares aujourd'hui. Certains auteurs ont ainsi proposé d'hybrider calculateurs photoniques et numériques (CPU/GPU) [25] quand d'autres développent des calculateurs purement analogiques pour l'IA [39]. Or, dès les débuts de la SAS, des implémentations matérielles temps-réel de l'ICA ont été proposées [29] et des travaux continuent en ce sens avec ces nouveaux hardwares [35]. De plus, ces nouveaux calculateurs pourraient amener à développer de nouveaux paradigmes pour l'apprentissage ou la SAS [61].

Changer de critères de performance

Compte tenu de la diversité des méthodes de SAS, il peut être intéressant de proposer de nouveaux critères de performance. En effet, on peut comparer des méthodes similaires en travaillant avec un budget de calcul donné [18], [60] ou comparer des méthodes d'apprentissages en mettant en regard leurs performances et leur complexité. Certains auteurs suggèrent ainsi d'intégrer d'autres mesures comme le coût carbone des calculs [31]. Mais comment comparer de manière équitable des méthodes statistiques non-supervisées, comme l'ICA, pouvant converger en quelques dizaines de secondes sur la CPU d'un ordinateur personnel, avec des approches d'apprentissage profond — probablement plus performantes au sens des mesures objectives utilisées en SAS [58] — mais ayant nécessité des journées d'entraînement sur des serveurs GPU dédiés ? Il nous semble nécessaire de réfléchir à de nouvelles mesures de performance objectives qui tiendraient compte du coût énergétique et environnemental.

Analyser le besoin et proposer des solutions adaptées

Comme nous l'avons indiqué précédemment, il n'est pas forcément nécessaire de développer des modèles coûteux pour résoudre une tâche donnée. La mission Planck de l'ESA nous permet d'illustrer ce propos : afin d'analyser le fond diffus cosmologique, c.-à-d. un rayonnement électromagnétique fossile datant d'environ 380 000 ans après le Big Bang, plusieurs équipes internationales ont proposé des approches non-supervisées ou supervisées de séparation de sources. Or, la carte de ce fond diffus cosmologique, qui a fait la une du New York Times en 2013, fut obtenue avec une approche aveugle d'ICA proposée par J.-F. Cardoso [3].

Plus récemment, nos travaux sur les signaux audios issus de boîtes noires aéronautiques [45] ont aussi cherché à ne pas aller vers des modèles coûteux. Lors d'un incident ou d'un accident, le Bureau d'enquêtes et d'analyse pour la sécurité de l'aviation civile (BEA) exploite les données issues des deux boîtes noires des avions pour en comprendre les causes. Une de ces boîtes est dédiée à l'ensemble des activités sonores captées dans le cockpit (d'où son nom de CVR, pour *cockpit voice recorder*) et est transcrite par un analyste audio pour les besoins de l'enquête de sécurité. Or, lors d'un incident, la superposition de nombreuses sources sonores (parole des pilotes, radio, alarmes, voix robotique du cockpit) dans le CVR rend très difficile cette tâche. La SAS apparaît alors comme un outil intéressant pour faciliter l'enquête des analystes. Cependant, le BEA ne souhaite pas utiliser des approches d'apprentissage profond dont les performances dépendent de la qualité de la phase d'entraînement. Par ailleurs, ces méthodes modernes ont plutôt cherché à rendre les sorties de SAS plus plaisantes à l'écoute, ce qui n'est pas forcément nécessaire pour

transcrire des signaux de parole. Aussi, nous avons plutôt cherché à reconstruire le modèle de mélange par rétro-ingénierie du CVR avant de mesurer la qualité de l'apport que des approches classiques de SAS (méthodes parcimonieuses et NMF) peuvent fournir à l'analyste audio [45]. De manière intéressante, on note que les approches testées ne fournissent pas les mêmes performances en fonction des mélanges considérés (mélange pilote/pilote vs mélange pilote/radio) et qu'en permettant à l'analyste d'écouter les sorties fournies par plusieurs méthodes de SAS (chacune ne nécessitant que quelques dizaines de secondes de calculs sur un ordinateur personnel), ce dernier réussit à transcrire une proportion significative de segments de parole initialement inintelligibles (44% des segments pilote/pilote et 66% des segments pilote/radio). Ces premiers résultats nous poussent à développer de nouvelles approches qui soient plus performantes tout en restant utilisables sur un simple ordinateur personnel.

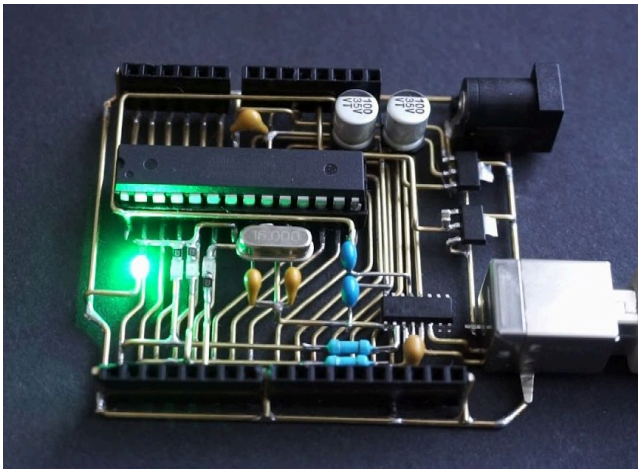

Fig. 3. Carte Freeduino de Jiří Praus (source [43]).

S'engager dans une démarche low-tech

La low-tech désigne une catégorie de techniques qui combinent durabilité, simplicité, résilience et facilité d'appropriation [9]. Cette catégorie semble donc complètement antagoniste avec la SAS ou toute autre discipline informatique. En pratique, on peut cependant se questionner sur le périmètre de la low-tech. Est-ce qu'une carte Arduino Uno facilement clonable (voir figure 3) — puisque ses plans et ses codes sont libres, et puisque les pièces nécessaires à sa fabrication sont disponibles et récupérables en très grande quantité — ne

pourrait pas être vue comme une centrale d'acquisition low-tech ? Pour les mêmes raisons, ne pourrait-on pas considérer un (rack de) Raspberry Pi comme une version low-tech de super-calculateur ? L'énergie pour alimenter un tel système (plusieurs Raspberry Pi et cartes Arduino) est sans commune mesure avec un *data center* moderne. Dans le cadre d'un calculateur universitaire low-tech, on pourrait imaginer un système alimenté par énergie renouvelable (voire boosté 1 h par jour, par exemple en faisant pédaler des étudiants en STAPS sur des vélos semblables à ceux utilisés pour recharger des smartphones dans les lieux publics). Cependant, même s'il est possible de déployer des modèles préentraînés sur de tels systèmes [15], à quel point est-il possible de réaliser aussi l'entraînement avec sur des données spécifiques ? Et quels volumes de données pourraient être stockés sur de tels systèmes low-cost ? Alors que nous empruntons une trajectoire opposée, et compte tenu des contraintes de ressources qui pèsent sur notre planète, c'est une voie qui nous semble pertinente à explorer.

Références

[1] Kinan Abbas, Matthieu Puigt, Gilles Delmaire, and Gilles Roussel. 2024. Locally-rank-one-based joint unmixing and demosaicing methods for snapshot spectral images. Part i: A matrix-completion framework. *IEEE Trans. Comput. Imaging*, vol. 10, p. 848–862.

[2] F. Abrard and Y. Deville. 2005. A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources. *Signal processing* 85, 7: 1389–1403.

[3] Yashar Akrami, M Ashdown, Jonathan Aumont, Carlo Baccigalupi, M Ballardini, Anthony J Banday, RB Barreiro, Nicola Bartolo, S Basak, K Benabed, and others. 2020. Planck 2018 results-IV. Diffuse component separation. *Astronomy & Astrophysics* 641: A4.

[4] Simon Arberet, Rémi Gribonval, and Frédéric Bimbot. 2010. A robust method to count and locate audio sources in a multichannel underdetermined mixture. *IEEE Trans. Signal Process.* 58, 1: 121–133.

[5] Richard G Baraniuk. 2011. More is less: Signal processing and the data deluge. *Science* 331, 6018: 717–719.

[6] Nematollah K Batmanghelich, Ben Taskar, and Christos Davatzikos. 2011. Generative-discriminative basis learning for medical imaging. *IEEE Trans. Med. Imag.* 31, 1: 51–69.

[7] Adel Belouchrani and Moeness G Amin. 1998. Blind source separation based on time-frequency signal representations. *IEEE Trans. Signal Process.* 46, 11: 2888–2897.

[8] Jignesh S Bhatt and Manjunath V Joshi. 2020. Deep learning in hyperspectral unmixing: A review. In *Proc. IEEE IGARSS'20*, 2189–2192.

[9] P. Bihouix. 2021. *L'âge des low tech. Vers une civilisation techniquement soutenable*. Editions du Seuil.

[10] José M Bioucas-Dias, Antonio Plaza, Nicolas Dobigeon, Mario Parente, Qian Du, Paul Gader, and Jocelyn Chanussot. 2012. Hyperspectral unmixing overview: Geometrical,

statistical, and sparse regression-based approaches. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 5, 2: 354–379.

[11] Chris Bissell. 2007. Historical perspectives - the moniac a hydromechanical analog computer of the 1950s. *IEEE Control Syst. Mag.* 27, 1: 69–74.

[12] Herbert Buchner, Robert Aichner, and Walter Kellermann. 2004. A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics. *IEEE transactions on speech and audio processing* 13, 1: 120–134.

[13] Dominique Cardon, Jean-Philippe Cointet, and Antoine Mazières. 2018. La revanche des neurones. *Réseaux* 211, 5: 173–220.

[14] I. Carron. The advanced matrix factorization jungle. (Dernier accès : 23/12/2022). `https://web.archive.org/web/20221205112424/https://sites.google.com/site/igorcarron2/matrixfactorizations`.

[15] Jiasi Chen and Xukan Ran. 2019. Deep learning with edge computing: A review. *Proc. IEEE* 107, 8: 1655–1674.

[16] Pierre Comon. 1994. Independent component analysis, a new concept? *Signal processing* 36, 3: 287–314.

[17] P. Comon and C. Jutten. 2010. *Handbook of blind source separation. Independent component analysis and applications*. Academic press.

[18] C. Dorffer, M. Puigt, G. Delmaire, and G. Roussel. 2017. Fast nonnegative matrix factorization and completion using Nesterov iterations. In *Proc. LVA/ICA'17*, 26–35.

[19] C. Dorffer, M. Puigt, G. Delmaire, and G. Roussel. 2018. Informed nonnegative matrix factorization methods for mobile sensor network calibration. *IEEE Trans. Signal Inf. Process. Netw.* 4, 4: 667–682.

[20] Michael Elad. 2010. *Sparse and redundant representations: From theory to applications in signal and image processing*. Springer Science & Business Media.

[21] H. Ferreboeuf, Z. Kahraman, M. Efoui-Hess, F. Berthoud, P. Bihouix, P. Fabre, D. Kaplan, L. Lefèvre, A. Monnin, O. Ridoux, and others. 2018. LEAN ICT–pour une sobriété numérique. *Rapport Intermédiaire du Groupe de Travail—The Shift Project; Agence française de développement and the Caisse des Dépôts: Paris, France*.

[22] Nicolas Gillis. 2020. *Nonnegative matrix factorization*. SIAM.

[23] Rémi Gribonval and Sylvain Lesage. 2006. A survey of sparse component analysis for blind source separation: Principles, perspectives, and new challenges. In *Proc. ESANN'06*, 323–330.

[24] Xiang Hao, Jibin Wu, Jianwei Yu, Chenglin Xu, and Kay Chen Tan. 2023. Typing to listen at the cocktail party: Text-guided target speaker extraction. *arXiv preprint arXiv:2310.07284*.

[25] D. Hesslow, A. Cappelli, I. Carron, L. Daudet, R. Lafargue, K. Müller, R. Ohana, G. Pariente, and I. Poli. 2021. Photonic co-processors in HPC: Using LightOn OPUs for randomized numerical linear algebra. `https://arxiv.org/abs/2104.14429`

[26] Aapo Hyvärinen and Petteri Pajunen. 1999. Nonlinear independent component analysis: Existence and uniqueness results. *Neural networks* 12, 3: 429–439.

[27] P. Jacquet, R. Rouvoy, and T. Ledoux. 2023. La chasse au gaspillage dans le cloud et les data centers.

[28] Christian Jutten and Pierre Comon. 2023. De la séparation de sources à l'analyse en composantes indépendantes, et au-delà. (Version longue). `https://hal.science/hal-04106245`

[29] Christian Jutten and Jeanny Hérault. 1990. Analog implementation of a permanent unsupervised learning algorithm. In *Neurocomputing: Algorithms, architectures and applications*. Springer, 145–152.

[30] Y. Koren, R. Bell, and C. Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8: 30–37.

[31] Alexandre Lacoste, Alexandra Luccioni, Victor Schmidt, and Thomas Dandres. 2019. Quantifying the carbon emissions of machine learning. *arXiv preprint arXiv:1910.09700*.

[32] D. D. Lee and H. S. Seung. 1999. Learning the parts of objects by non negative matrix factorization. *Nature* 401, 6755: 788–791.

[33] Pengfei Li, Jianyi Yang, Mohammad A Islam, and Shaolei Ren. 2023. Making AI less" thirsty": Uncovering and addressing the secret water footprint of ai models. *arXiv preprint arXiv:2304.03271*.

[34] Chao Liu, Hung-chih Yang, Jinliang Fan, Li-Wei He, and Yi-Min Wang. 2010. Distributed nonnegative matrix factorization for web-scale dyadic data analysis on MapReduce. In *Proceedings of the 19th international conference on world wide web*, 681–690.

[35] Philip Y Ma, Alexander N Tait, Weipeng Zhang, Emir Ali Karahan, Thomas Ferreira De Lima, Chaoran Huang, Bhavin J Shastri, and Paul R Prucnal. 2020. Blind source separation with integrated photonics and reduced dimensional statistics. *Optics letters* 45, 23: 6494–6497.

[36] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. 2009. Online dictionary learning for sparse coding. In *Proc. ICML*, 689–696.

[37] Vishal Monga, Yuelong Li, and Yonina C Eldar. 2021. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Process. Mag.* 38, 2: 18–44.

[38] Aditya Arie Nugraha, Antoine Liutkus, and Emmanuel Vincent. 2016. Multichannel audio source separation with deep neural networks. *IEEE/ACM Trans. Audio, Speech, Language Process.* 24, 9: 1652–1664.

[39] Marcus Valtonen Örnhag, Püren Güler, Dmitry Knyaginin, and Mattias Borg. 2023. Accelerating AI using next-generation hardware: Possibilities and challenges with analog in-memory computing. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 488–496.

[40] Alexey Ozerov and Cédric Févotte. 2009. Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation. *IEEE Audio, Speech, Language Process.* 18, 3: 550–563.

[41] Lucas Parra and Clay Spence. 2000. Convolutive blind separation of non-stationary sources. *IEEE Trans. Speech Audio Process.* 8, 3: 320–327.

[42] D. Patterson, J. Gonzalez, Q. Le, C. Liang, L.-M. Munguia, D. Rothchild, D. So, M. Texier, and J. Dean. 2021. Carbon emissions and large neural network training. *arXiv preprint arXiv:2104.10350*.

[43] Jiří Praus. 2019. A skeleton arduino uno. `https://www.hackster.io/jiripraus/a-skeleton-arduino-uno-89bdd6`.

[44] M. Puigt. 2007. Méthodes de séparation aveugle de sources fondées sur des transformées temps-fréquence. Application à des signaux de parole. Université Toulouse III – Paul Sabatier.

[45] M. Puigt, B. Bigot, and H. Devulder. 2025. Introducing the “cockpit party problem”: Blind source separation enhances aircraft cockpit speech transcription. _Journal of the Audio Engineering Society 73, 1/2.

[46] M. Puigt, A. Griffin, and A. Mouchtaris. 2011. Post-nonlinear speech mixture identification using single-source temporal zones & curve clustering. In *Proc. EUSIPCO 2011*, 1844–1848.

[47] M. Puigt, A. Griffin, and A. Mouchtaris. 2012. Nonlinear blind mixture identification using local source sparsity and functional data clustering. In *Proc. IEEE SAM 2012*, 489–492.

[48] Hendrik Purwins, Bo Li, Tuomas Virtanen, Jan Schlüter, Shuo-Yiin Chang, and Tara Sainath. 2019. Deep learning for audio signal processing. *IEEE J. Sel. Topics Signal Process.* 13, 2: 206–219.

[49] Semiconductor Industry Association and Semiconductor Research Corporation. 2015. Rebooting the IT revolution: A call to action.

[50] Paris Smaragdis, Bhiksha Raj, and Madhusudana Shashanka. 2007. Supervised and semi-supervised separation of sounds from single-channel mixtures. In *Proc. ICA'07*, 414–421.

[51] A. Stéphant. 2022. Transition énergétique : Une nécessaire intégration des impacts environnementaux de l'industrie minière. *Revue internationale et stratégique*, 128: 95–103.

[52] Thomas N Theis and H-S Philip Wong. 2017. The end of Moore's law: A new beginning for information technology. *IEEE Comput. Sci. Eng.* 19, 2: 41–50.

[53] Y. Tsividis. 2018. Not your father's analog computer. *IEEE Spectr.* 55, 2: 38–43.

[54] Laurens van Der Maaten, Eric Postma, and Jaap van den Herik. 2009. Dimensionality reduction: A comparative review. *J Mach Learn Res* 10, 66-71: 13.

[55] René Vidal. 2011. Subspace clustering. *IEEE Signal Process. Mag.* 28, 2: 52–68.

[56] E. Vincent. 2007. Complex nonconvex lp norm minimization for underdetermined source separation. In *Proc. ICA 2007*, 430–437.

[57] Emmanuel Vincent, Shoko Araki, Fabian Theis, Guido Nolte, Pau Bofill, Hiroshi Sawada, Alexey Ozerov, Vikrham Gowreesunker, Dominik Lutter, and Ngoc QK Duong. 2012. The signal separation evaluation campaign (2007–2010): Achievements and remaining challenges. *Signal Processing* 92, 8: 1928–1936.

[58] Emmanuel Vincent, Rémi Gribonval, and Cédric Févotte. 2006. Performance measurement in blind audio source separation. *IEEE Audio, Speech, Language Process.* 14, 4: 1462–1469.

[59] DeLiang Wang and Jitong Chen. 2018. Supervised speech separation based on deep learning: An overview. *IEEE/ACM Trans. Audio, Speech, Language Process.* 26, 10: 1702–1726.

[60] F. Yahaya, M. Puigt, G. Delmaire, and G. Roussel. 2024, A framework for compressed weighted nonnegative matrix factorization, IEEE Trans. on Signal Process. 72, 4798-4811. https://dx.doi.org/10.1109/TSP.2024.3469830.

[61] F. Yahaya, M. Puigt, G. Delmaire, and G. Roussel. 2021. Random projection streams for (weighted) nonnegative matrix factorization. In *Proc. IEEE ICASSP'21*.

[62] Ozgur Yilmaz and Scott Rickard. 2004. Blind separation of speech mixtures via time-frequency masking. *IEEE Trans. Signal Process.* 52, 7: 1830–1847.

[63] Guoxu Zhou, Andrzej Cichocki, Qibin Zhao, and Shengli Xie. 2014. Nonnegative matrix and tensor factorizations: An algorithmic perspective. *IEEE Signal Process. Mag.* 31, 3: 54–65.