

**Université de Montréal**

**Accélération du lentillage gravitationnel à plans multiples par  
apprentissage profond**

par

**Charles Wilson**

Département de physique  
Faculté des arts et des sciences

Mémoire présenté en vue de l'obtention du grade de  
Maître ès sciences (M.Sc.)  
en physique

Avril 2023



**Université de Montréal**

Faculté des arts et des sciences

Ce mémoire intitulé

**Accélération du lentillage gravitationnel à  
plans multiples par apprentissage profond**

présenté par

**Charles Wilson**

a été évalué par un jury composé des personnes suivantes :

*Laurence Perreault-Levasseur*

(président-rapporteur)

*Yashar Hezaveh*

(directeur de recherche)

*Björn Benneke*

(membre du jury)



## Résumé

---

Le "modèle standard" actuel de la cosmologie est celui de  $\Lambda$ CDM, décrivant un Univers en expansion accélérée ainsi qu'une structure de matière sombre froide formée en halos, sur lesquels s'assemblent les galaxies. Malgré les nombreuses confirmations observationnelles de ses prédictions, il existe d'importantes tensions entre les mesures de la distribution de structure sombre aux petites échelles de l'Univers et ce qui serait attendu de  $\Lambda$ CDM. Cependant, ces halos légers de matière sombre, qui sont prédit d'abonder à travers le cosmos, n'hébergent pas de galaxies lumineuses et sont donc très difficiles à observer directement. Leur présence peut toutefois être détectée dans les lentilles gravitationnelles fortes de type galaxie-galaxie, un phénomène se produisant lorsque la lumière d'une galaxie d'arrière-plan est fortement déviée par le champ gravitationnel d'une galaxie d'avant-plan, formant des images multiples et des arcs étendus. Les halos distribués en ligne de visée de tels systèmes, ainsi que ceux imbriqués dans la galaxie lentille, peuvent causer des perturbations gravitationnelles dans les images de galaxies lentillées. La détection de ces effets infimes dans des observations de lentilles gravitationnelles est faite par des méthodes statistiques Bayésiennes, qui nécessitent des centaines de milliers de simulations de la contribution de ces perturbateurs à la déflexion de la lumière. Traditionnellement, la modélisation du lentillage par les halos en ligne de visée s'est faite avec le formalisme du lentillage à plans multiples, qui souffre d'une nature récursive peu efficace. De plus, il est prédit par le modèle  $\Lambda$ CDM que la majorité des systèmes de lentilles gravitationnelles comporteraient davantage de halos en ligne de visée que de sous-halos imbriqués dans la galaxie lentille, motivant une modélisation détaillée des effets de ligne de visée. Dans un contexte d'analyse Bayésienne, l'approche du lentillage à plans multiples représente une échelle de temps de plusieurs jours pour l'analyse d'un seul système. En considérant que des grands relevés du ciel comme ceux de l'Observatoire Vera Rubin et du télescope spatial Euclid sont projetés de découvrir des centaines de milliers de lentilles gravitationnelles, l'effort de contraindre la distribution de matière sombre aux petites échelles se voit confronté à ce qui pourrait être un insurmontable problème de temps de calcul.

Dans ce mémoire, je présente le développement d'un nouveau formalisme de modélisation du lentillage gravitationnel par halos en ligne de visée accéléré par des réseaux de neurones, motivé par les lacunes du lentillage à plans multiples et l'importance scientifique de la modélisation de ces effets. Les architectures de ces réseaux, conçues dans le cadre de ce travail, sont basées sur le mécanisme d'attention, et peuvent être conditionnées sur des ensembles de modèles de halos en ligne de visée afin de produire les angles de déflexion leur étant associés. Ce formalisme offre la flexibilité requise pour remplacer celui du lentillage à plans multiples, laissant à l'utilisateur la liberté de spécifier un modèle de lentille principale et étant compatible avec des grilles de pixels de taille quelconque. Notre formalisme permet d'accélérer la modélisation du lentillage de ligne de visée par presque deux ordres de grandeur lorsque comparé au lentillage à plans multiples, et promet d'atteindre une exactitude lui étant comparable dans des développements futurs. Il s'agit d'une contribution significative à l'étude de la matière sombre aux petites échelles, qui permettra soit de réconcilier  $\Lambda$ CDM et les observations, ou mènera à l'adoption d'un modèle cosmologique alternatif.

**Mots-clés:** Lentilles gravitationnelles — Matière sombre — Cosmologie — Simulations astrophysiques — Réseaux de neurones

## Abstract

---

The current "standard model" of cosmology is that of  $\Lambda$ CDM, describing a Universe undergoing accelerated expansion with a structure of cold dark matter formed into halos, onto which are assembled galaxies. Despite the numerous observational confirmations of its predictions, there remains some important tensions between measures of the distribution of dark structure on small scales of the Universe and what would be expected from  $\Lambda$ CDM. However, these light dark matter halos, predicted to be abundant throughout the cosmos, are not hosts of luminous galaxies and are therefore very difficult to observe directly. Yet, their presence can still be detected in galaxy-galaxy type strong gravitational lenses, a phenomenon occurring when the light of a background galaxy is strongly deflected by the gravitational field of a foreground galaxy, forming multiple images and extended arcs. Halos distributed along the line-of-sight of such systems, as well as those nested within the lens galaxy, can introduce gravitational perturbations in images of lensed galaxies. The detection of such infinitesimal effects in strong lensing observations is made with methods relying on Bayesian statistics, which require hundreds of thousands of simulations of the contribution of these perturbers to the deflection of light. Traditionally, modeling the lensing from line-of-sight halos has been done with the multi-plane lensing framework, which suffers from its inefficient recursive nature. Moreover, the  $\Lambda$ CDM model predicts that most gravitational lens systems would host a larger amount of line-of-sight halos than subhalos nested within the lens galaxy, motivating a detailed modeling of line-of-sight effects. In a Bayesian analysis context, the multi-plane lensing approach represents a timescale of multiple days for the analysis of a single system. Considering that large sky surveys such as those of the Vera Rubin Observatory and the Euclid space telescope are projected to discover hundreds of thousands of gravitational lenses, the effort of constraining the small-scale distribution of dark matter is confronted to what might seem like an insurmountable problem of computation time.

In this thesis, I present the development of a new neural-network-accelerated framework for modeling the gravitational lensing by line-of-sight halos, motivated by the shortcomings of multi-plane lensing and the scientific importance of modeling these effects. The architectures of these networks, conceived as part of this work, are based on the attention mechanism, and can be conditioned on sets of line-of-sight halo models in order to produce their associated deflection angles. This framework offers the flexibility required to replace that of multi-plane lensing, leaving up to the user the freedom to specify a main lens model and being compatible with pixel grids of any size. Our framework allows to accelerate the modeling of line-of-sight lensing by nearly two orders of magnitude relative to multi-plane lensing, and promises to reach a comparable accuracy in future developments. This constitutes a significant contribution to the study of dark matter on small scales, which will either lead to the reconciliation of  $\Lambda$ CDM and observations, or the adoption of an alternate cosmological model.

**Keywords:** Strong gravitational lensing — Dark matter — Cosmology — Astronomical simulations — Neural networks



# Table des matières

---

<b>Résumé</b> .....	v
<b>Abstract</b> .....	vii
<b>Liste des tableaux</b> .....	xiii
<b>Liste des figures</b> .....	xv
<b>Liste des abréviations</b> .....	xix
<b>Liste des symboles</b> .....	xxi
<b>Remerciements</b> .....	xxvii
<b>Chapitre 1. Introduction</b> .....	1
1.1. Description du mémoire.....	5
1.2. Déclaration de l'étudiant.....	5
<b>Chapitre 2. Matière sombre</b> .....	7
2.1. Historique.....	7
2.2. La cosmologie moderne en crise.....	12
<b>Chapitre 3. Lentillage gravitationnel fort</b> .....	17
3.1. Principe de Fermat et angles de déflexion.....	18
3.2. Équation de lentille.....	22
3.2.1. Traçage de rayons inverse.....	24
3.3. Lentillage gravitationnel à plans multiples.....	25
<b>Chapitre 4. Apprentissage automatique</b> .....	29

4.1.	Fondements de l'apprentissage automatique .....	30
4.1.1.	Modèles linéaires .....	30
4.1.2.	Réseaux de neurones à propagation avant .....	31
4.2.	Entraînement d'un modèle .....	35
4.2.1.	Rétropropagation des gradients .....	35
4.2.2.	Algorithmes d'optimisation .....	38
4.3.	Réseaux autoattentionnels .....	40
4.3.1.	Mécanisme d'attention .....	40
4.3.2.	Réseaux autoattentionnels d'ensembles .....	41
<b>Chapitre 5.</b>	<b>AlphaGen: a Neural Network for Fast Multi-Plane Lensing .....</b>	<b>45</b>
	Résumé .....	46
	Abstract .....	46
5.1.	Introduction .....	47
5.2.	Lensing formalism .....	50
5.2.1.	Lens equation and lensing quantities .....	50
5.2.2.	Multiplane lensing .....	51
5.3.	The AlphaGen Pipeline .....	52
5.3.1.	Set Model .....	53
5.3.2.	Map Model .....	55
5.3.3.	Objective function .....	55
5.3.4.	Generation pipeline .....	55
5.4.	Data .....	58
5.4.1.	Halo mass profile .....	58
5.4.2.	Sampling halo populations .....	59
5.4.3.	Producing datasets .....	60
5.5.	Training .....	61
5.5.1.	Feature normalization .....	61
5.5.2.	Foreground Model .....	62

5.5.3. Background Model .....	63
5.6. Results and Discussion .....	63
5.6.1. Test set performance .....	63
5.6.2. Lensing performance .....	66
5.6.3. Acceleration .....	69
5.7. Summary and Conclusion .....	70
Software .....	71
Acknowledgements .....	71
<b>Chapitre 6. Conclusion .....</b>	<b>73</b>
<b>Références bibliographiques .....</b>	<b>77</b>
<b>Annexe A. <math>\Lambda</math>CDM .....</b>	<b>97</b>



## Liste des tableaux

---

5.1	Detailed architecture of the Foreground and Background models. ....	62
5.2	Training hyperparameters of the Foreground and Background models .....	62
5.3	Main deflector model parameters.....	63
A.1	Paramètres du modèle $\Lambda$ CDM mesurés à partir d’observations du fond diffus cosmologique prises par le télescope Planck (Planck Collaboration <i>et al.</i> , 2020).....	97



## Liste des figures

---

- 1.1 **Gauche:** simulation du lentillage gravitationnel fort d'une galaxie source lointaine par le halo de matière sombre d'une galaxie lentille (GL) entre la source et l'observateur. **Centre:** lentillage de la même source par le même modèle de galaxie lentille, cette fois-ci incluant les effets d'une population de sous-halos dans le plan lentille et de halos distribués en ligne de visée. **Droite:** différence entre les deux simulations de lentille. Image de galaxie source provenant du recensement COSMOS (crédit: ESA/Hubble & NASA). . . . . 2
- 1.2 Schéma d'une population de halos légers dans un certain système de lentille gravitationnelle. La source se situe du côté gauche, émettant des rayons de lumière représentés par les traits rouges. Les trajectoires de ceux-ci sont principalement déviées par la galaxie lentille se trouvant dans le disque central, mais aussi par une multitude d'autres halos, avant de converger vers l'observateur du côté droit. Des halos de champ, distribués le long de la direction de décalage vers le rouge (*redshift*), sont représentés par des cercles, comme le sont les sous-halos dans le plan de la lentille. La taille et la couleur des cercles varie de façon proportionnelle à la masse et la concentration des halos, respectivement. Figure tirée de Gilman *et al.* (2022). . . . . 3
- 2.1 Vitesses orbitales moyennes dans le plan galactique en fonction de la distance radiale du noyau pour un échantillon de 21 galaxies spirales de type Sc, arrangées en ordre croissant de rayon galactique. On y remarque que les courbes de vitesse deviennent "plates" au-delà de la région du bulbe central. Figure tirée de Rubin *et al.* (1980). . . . . 10
- 2.2 Image composite de l'amas de la Balle, montrant le résultat d'une collision frontale entre deux amas de galaxies. Le gaz chaud de matière baryonique des amas est représenté en rose, imagé en rayons X par le télescope Chandra. Les halos de matière sombre de chaque amas, détectés par le lentillage gravitationnel faible d'objets d'arrière-plan, sont représentés en bleu. L'image dans le visible des galaxies de

	champ et membres de l'amas provient des télescopes Magellan et Hubble. Crédit: rayons X: NASA/CXC/CfA/Markevitch <i>et al.</i> (2002); lentillage: NASA/STScI, ESO WFI, Magellan/U.Arizona/Clowe <i>et al.</i> (2006); visible: NASA/STScI, Magellan/U.Arizona/Clowe <i>et al.</i> (2006).....	10
2.3	<b>Gauche:</b> réalisation de sous-structure dans un halo massif comparable à celui de la Voie Lactée ( $\sim 10^{12} M_{\odot}$ ), selon une fonction de masse correspondant à CDM. <b>Droite:</b> réalisation de sous-structure dans le même halo, cette fois selon une fonction de masse de WDM ayant un point tournant de troncation vers $10^8 M_{\odot}$ . Crédit: Gilman (2020) ..	14
3.1	Schéma d'une lentille gravitationnelle galaxie-galaxie dans l'approximation de la lentille mince. Le trait noir épais représente la trajectoire d'un photon qui se fait dévier par la lentille du plan en $z_{\ell}$ .....	24
3.2	SDSS J1148+1930, aussi appelée le fer à cheval cosmique, est un exemple impressionnant d'une lentille gravitationnelle galaxie-galaxie. La distante galaxie bleue à $z_s = 2.381$ apparaît en plusieurs images fortement distordues, formant un anneau presque parfait d'un énorme rayon d'Einstein d'environ $5''$ . La galaxie rouge à $z_{\ell} = 0.444$ est responsable de cet effet de lentillage; avec une masse estimée de $5.2 \times 10^{12} M_{\odot}$ (Schuldt <i>et al.</i> , 2019), il s'agirait de l'une des galaxies les plus massives jamais observées. Composite d'images en visible et infrarouge prises par l'instrument WFC3 de HST. Crédit: ESA/Hubble & NASA.....	25
3.3	Schéma du lentillage à plans multiples pour une lentille gravitationnelle galaxie-galaxie avec de la structure en ligne de visée. Le trait noir épais représente la trajectoire d'un photon. ....	27
4.1	Schéma d'un réseau de neurones à trois couches. Dans ce choix de nomenclature, "trois couches" fait référence aux deux couches cachées avec paramètres $\mathbf{W}_{(1)}$ et $\mathbf{W}_{(2)}$ ainsi qu'à la couche de sortie $\mathbf{W}_{(3)}$ . La couche d'entrée n'est pas comptée dans le nombre de couches.....	32
4.2	Graphe de calcul associé au réseau de neurone de la figure 4.1. Les vertex impliqués dans le gradient de $\mathcal{L}$ p.r.à $\mathbf{W}_{(2)}$ sont identifiés en bleu.....	35



4.3	<p>Courbes de niveau de la fonction objectif (4.1.2) et trajectoires de différents algorithmes d’optimisation dans l’espace des paramètres à deux dimensions d’un problème de régression linéaire. Chacun des panneaux illustre 20 pas d’entraînement à partir des mêmes paramètres initiaux, pour des mini-lots de 2 exemples. Les courbes de niveaux de la fonction objectif <math>\mathcal{L}</math> correspondent aux valeurs de <math>\mathcal{L}</math> évaluée pour l’ensemble de données entier de 100 exemples. ....</p>	39
5.1	<p>Multi-plane lensing, shown here with a single deflector per lens plane. The dotted horizontal line is the optical axis, and its intersection with each of the <math>N + 1</math> lens planes is the origin for their angular coordinates <math>\theta_i</math>. The thick solid line represents the trajectory of a light ray. In backward ray-tracing, the ray comes in through the lens plane at <math>z_\ell</math> at the angular position <math>\theta_\ell</math>, and is deflected by an angle <math>\hat{\alpha}_\ell</math> relative to the plane. This deflection determines the angular position <math>\theta_{\ell+1}</math> at the next lens plane through eq. (5.2.6).</p>	52
5.2	<p>Diagram of our network architecture for the Foreground Model. The network shown on top is a Set Model, which operates layers of self-attention and pooling on an input set of halo parameters. On the bottom is a Map Model, receiving an encoding of the halo population along with initial ray angles at <math>z_o</math> and a conditioned redshift <math>z</math> to produce effective deflection angles at said redshift. Shown below the <math>1 \times 1</math> convolution blocks of the Map Model are the number of filters, and along their depth is the side dimension of the square feature maps. The Background Model’s architecture is identical to this except for two additional channels in the Map Model input for the outgoing ray angles <math>\hat{\alpha}_\ell</math> at <math>z_\ell</math>.</p>	56
5.3	<p>Foreground (left column) effective deflection angles <math>\alpha_{\text{eff}}</math> and background (right column) halo contribution <math>\Delta\alpha_{\text{bg}}</math> to total deflection angles associated to four test halo populations and main deflector models (along rows), comparing angles computed from true MPRT and from the Foreground and Background models of the AlphaGen pipeline. The background angles are typically of a higher amplitude than those of the foreground, due to the larger comoving volume and number of halos. Reconstructions of angles at a pixel size of <math>128 \times 128</math>, corresponding to the size of these models’ training labels. ....</p>	64
5.4	<p>Example of the reconstruction of foreground angles associated to a test halo population of a different size than the training halo populations. While the intricate spatial distribution of the angles is well reproduced, there is a manifest directional bias in the residuals. ...</p>	66

5.5	Simulated lensing of COSMOS sources of normalized intensity including deflections from LOS halos, with effective multi-plane deflection angles computed with true MPRT (first column) and the AlphaGen pipeline (second column). The residuals between these two models (third column) are expressed in terms of a noise level $\sigma = 0.05$ , so that the pSNR the observations would be $\sim 1/\sigma$ . Also shown are lensed COSMOS sources modeled with the LOS projection scheme (fourth column) and their associated residuals (fifth column) with respect to the true MPRT lenses.....	68
5.6	(Top) Runtime of the generation of multi-plane deflection angles as a function of the number of pixels in the field of view, for the true MPRT computation (red) and the forward passes of our pipeline's networks (blue). For each pixel size, 10 generations of test halo populations of variable size were sampled and the production of their deflection angles was timed for both methods. The points above mark the mean runtimes across these 10 generations, and barely visible errorbars correspond to the standard deviation of the runtimes. Note that the time axis is in log-scale. (Bottom) Ratio of the mean runtimes between true MPRT and our pipeline.....	70

## Liste des abréviations

---

CDM	Cold Dark Matter — Matière sombre froide
WDM	Warm Dark Matter — Matière sombre tiède
SIDM	Self-Interacting Dark Matter — Matière sombre auto-interactive
GL	Galaxie lentille
MOND	Modified Newtonian Dynamics — Dynamique newtonienne modifiée
TeV <sub>S</sub>	Tensor-vector-scalar gravity — gravité tenseur-vecteur-scalaire
CMB	Cosmic Microwave Background — Fond diffus cosmologique
LOS	Line-of-sight — Ligne de visée
MPRT	Multi-Plane Ray-Tracing — Traçage de rayons à plans multiples
FLRW	Friedmann-Lemaître-Robertson-Walker
IA	Intelligence Artificielle
ESA	European Space Agency
NASA	National Aeronautics and Space Administration
CXC	Chandra X-Ray Center
CfA	Center for Astrophysics   Harvard & Smithsonian
STScI	Space Telescope Science Institute
ESO WFI	European Southern Observatory - Wide Field Imager

HST	Hubble Space Telescope
WFC3	Wide Field Camera 3
RGB	Red Green Blue
SGD	Stochastic Gradient Descent — Descente par gradient stochastique
RAE	Réseau(x) Autoattentif(s) d'Ensemble
rFF	row Feed-Forward
FC	Fully-Connected — Couche linéaire
MAB	Multi-head Attention Block
SAB	Set Attention Block
ISAB	Induced Set Attention Block
PMA	Pooling by Multi-head Attention
MLP	Multi-Layer Perceptron — Perceptron multicouche
CNN	Convolutional Neural Network — Réseau de neurones convolutif
NFW	Profil de densité de Navarro, Frenk & White
SNR	Signal-to-Noise Ratio — Rapport signal sur bruit
AG	AlphaGen
c.-à-d.	c'est-à-dire
p.r.à/a.	par rapport à/aux

## Liste des symboles

---

$\Lambda$	Constante cosmologique
pc	Parsec
AU	Unité astronomique
as/ ”	Seconde d’arc
°	Degré
eV	Électronvolt
$\pi$	Pi
$M_{\odot}$	Masse du Soleil
$L_{\odot}$	Luminosité du Soleil
$\Upsilon$	Rapport masse/luminosité
$\theta_E$	Rayon d’Einstein
$G$	Constante universelle de gravitation
$c$	Vitesse de la lumière
$D_s$	Distance de diamètre angulaire de l’observateur à la source
$D_{\ell}$	Distance de diamètre angulaire de l’observateur à un plan de lentille
$D_{\ell s}$	Distance de diamètre angulaire d’un plan de lentille à la source
$z$	Décalage vers le rouge

$a$	Facteur d'échelle $a = (1 + z)^{-1}$
$\lambda_{\text{FS}}$	Échelle de dérive libre
$\epsilon_{\star}$	Efficacité de formation d'étoiles
$\mathbb{R}$	Ensemble des nombres réels
$\Phi$	Potentiel gravitationnel
$\lambda$	Paramètre de trajectoire d'un photon
$\mathbf{q}$	Coordonnée physique d'un photon - vecteur en $\mathbb{R}^3$
$\dot{\mathbf{q}}$	Dérivée de $\mathbf{q}$ p.r.à $\lambda$ - vecteur en $\mathbb{R}^3$
$g_{\mu\nu}$	Élément de la métrique $g$
$t$	Coordonnée de temps
$\tau$	Temps propre
$\mathcal{L}$	Lagrangien ou fonction d'erreur
$\nabla$	Gradient
$\nabla^2$	Laplacien
$\hat{\alpha}$	Angle de déflexion physique - vecteur en $\mathbb{R}^2$
$\alpha$	Angle de déflexion réduit - vecteur en $\mathbb{R}^2$
$\zeta$	Coordonnée comobile parallèle à la ligne de visée
$\xi$	Coordonnées comobiles dans le plan image - vecteur en $\mathbb{R}^2$
$\eta$	Coordonnées comobiles dans le plan source - vecteur en $\mathbb{R}^2$
$\theta$	Coordonnées angulaires dans le plan image - vecteur en $\mathbb{R}^2$
$\beta$	Coordonnées angulaires dans le plan source - vecteur en $\mathbb{R}^2$

$\rho$	Densité
$\Sigma$	Densité surfacique
$\Sigma_c$	Densité surfacique critique
$\kappa$	Convergence - densité surfacique en unités de densité critique
$\mathcal{O}(\cdot)$	Complexité de calcul
$\mathcal{D}$	Jeu de données
$\mathbf{x}^{(i)}$	$i$ -ème exemple (entrée) d'un jeu de données - vecteur en $\mathbb{R}^d$
$\mathbf{y}^{(i)}$	Cible associée à un exemple $\mathbf{x}^{(i)}$ - vecteur en $\mathbb{R}^n$
$\hat{\mathbf{y}}^{(i)}$	Prédiction d'un modèle ayant reçu en entrée un exemple $\mathbf{x}^{(i)}$ - vecteur en $\mathbb{R}^n$
$w$	Poids d'un modèle
$\mathbf{w}$	Ensemble des poids d'un modèle
$\mathbf{w}^*$	Ensemble des poids optimaux d'un modèle
$\phi$	Fonctions de base
$\ \cdot\ _2$	Norme euclidienne $\ell_2$
$h_k^{(i)}$	Préactivation du $k$ -ième neurone de la $i$ -ième couche d'un réseau de neurones
$a_k^{(i)}$	Activation du $k$ -ième neurone de la $i$ -ième couche d'un réseau de neurones
$\sigma(\cdot)$	Fonction d'activation non-linéaire dans une couche d'un réseau de neurones
$\eta$	Taux d'apprentissage
$\odot$	Produit de Hadamard
$\times_n$	Produit mode- $n$
$\mathbb{E}[\cdot]$	Opérateur d'espérance

$Q, K, V$	Matrices de requêtes, clés et valeurs dans une opération d'attention
$\omega$	Fonction d'activation dans une opération d'attention
$h$	Population de halos de champs
$n_{\text{pix}}$	Nombre de pixels
$\mathcal{U}(a, b)$	Distribution uniforme avec support dans l'intervalle $[a, b]$



À grand-maman



## Remerciements

---

Je tiens à remercier mon superviseur, Yashar Hezaveh, qui par sa nature curieuse et sa passion infectieuse pour la science, m'a poussé à me dépasser et à trouver ma voie en tant que scientifique.

Je tiens à remercier Alexandre Adam, Ève Campeau-Poirier, Ronan Legin, et tous les autres membres de l'Institut CIELA avec qui j'ai eu le bonheur de passer ces dernières années à parler de vieilles bêtises jusqu'aux trucs plus intelligents. Vous m'avez beaucoup appris et j'ai trouvé auprès de vous des véritables amis. Essayez de ne pas découvrir tous les secrets de l'Univers trop vite, laissez moi-en quelques-uns.

Merci aussi à Adam Coogan et Stéphanie Luna, qui ont été des mentors que j'admire beaucoup. Merci Stéphanie de m'avoir botté le cul pour commencer à écrire ce mémoire au bon moment.

Merci à Charles Gauthier, mon partenaire de lab, d'avoir partagé cette dernière ride avec moi pendant plus longtemps que prévu.

Je réserve également une pensée particulièrement spéciale à mes amis Joseph, William, Nicolas, Louis-Philippe et Oreste. Merci d'avoir vécu avec avec moi une histoire encore plus farfelue que celle que j'ai racontée dans ce manuscrit pendant ces deux dernières années, et d'être une des meilleures raisons d'habiter à Montréal.

Enfin, je remercie mes parents, Sophie et Rénaud, pour leur soutien inconditionnel et leur amour pendant ces dernières années et toutes celles qui sont venues avant. Merci à mes frères, je vous aime beaucoup. Mathias, je ne suis désormais plus "dans ma maîtrise".



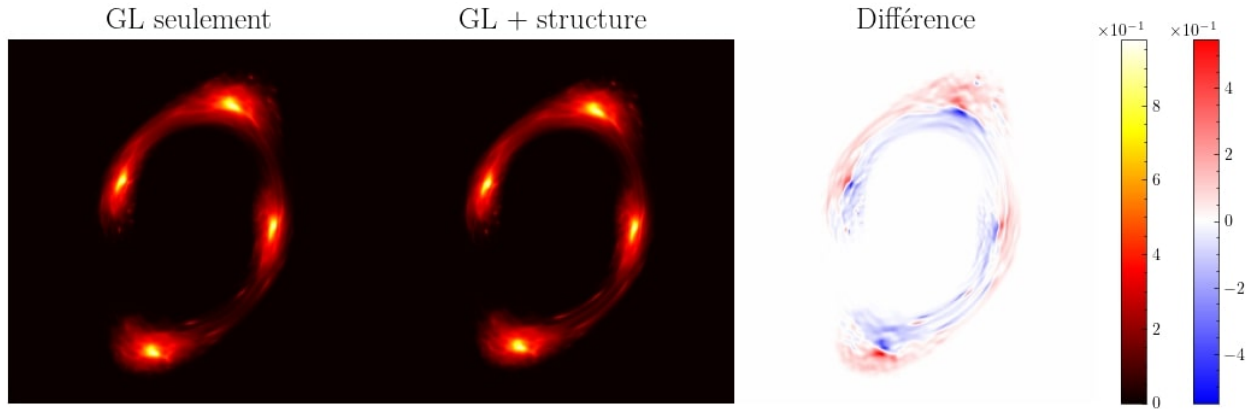
# Chapitre 1

---

## Introduction

En cette ère de la cosmologie de précision, une multitude de mesures astrophysiques sont venues appuyer les prédictions du modèle cosmologique de concordance,  $\Lambda$ CDM. Or, la plupart de ses succès soulèvent à leur tour d'importantes questions auxquelles le Modèle Standard de la physique des particules n'offre toujours pas de réponses. Notamment, quelle est la nature particulière et l'origine de la matière sombre, cette composante inconnue n'interagissant pas avec la force électromagnétique qui compterait pour 80% de la masse dans l'Univers? Cette dernière est essentielle à notre compréhension du fond diffus cosmologique, et est centrale aux arguments de formation de structure de  $\Lambda$ CDM (Planck Collaboration *et al.*, 2020). Ceux-ci stipulent que des *halos*, soit des nuages de matière sombre liés gravitationnellement, forment le squelette cosmique sur lequel naissent les galaxies et les amas de galaxies. La distribution de cette structure de matière sombre est très bien contrainte par les observations aux plus grandes échelles de l'Univers, c.-à-d. au-delà de l'ordre de  $\sim 1$  Mpc, où elle tisse une toile cosmique sous la forme d'immenses filaments galactiques séparés de grands vides. Toutefois, c'est aux échelles sous-galactiques que les mesures et les prédictions de  $\Lambda$ CDM semblent être en désaccord (Bullock et Boylan-Kolchin, 2017).

Selon ce modèle, d'innombrables halos de matière sombre aux plus petites échelles de masse parsèment le cosmos, la plupart de ceux-ci étant trop légers pour abriter des galaxies. Cependant, seul un infime nombre de ces halos légers ont été détectés jusqu'à présent. Cela impliquerait que ces halos dépourvus de matière lumineuse sont soit trop obscurs pour être détectés directement, soit ils n'existent tout simplement pas. Certains modèles alternatifs à  $\Lambda$ CDM prédisent que les halos ne se forment pas sous une certaine échelle de masse, échelle qui dépend des propriétés particulières de la matière sombre. Ainsi, une non-détection de cette population de halo légers invaliderait  $\Lambda$ CDM, favorisant ces autres modèles et renversant le paradigme cosmologique actuel. Alternativement, la confirmation de l'existence de ces halos irait à appuyer l'une des prédictions principales de  $\Lambda$ CDM.

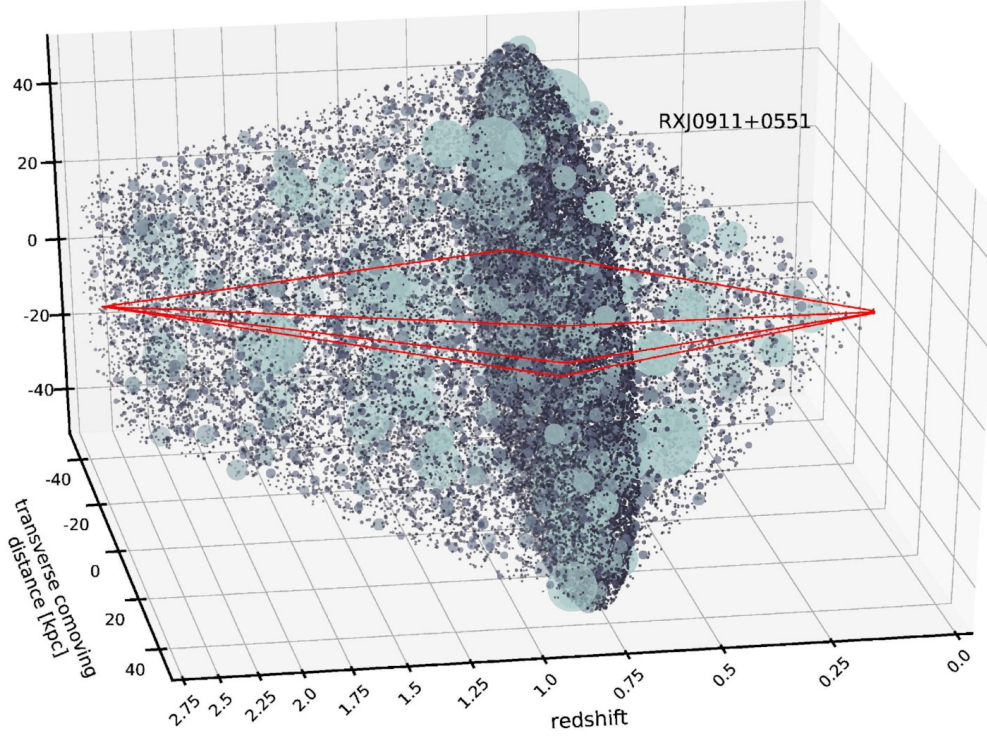


**Fig. 1.1. Gauche:** simulation du lentillage gravitationnel fort d’une galaxie source lointaine par le halo de matière sombre d’une galaxie lentille (GL) entre la source et l’observateur. **Centre:** lentillage de la même source par le même modèle de galaxie lentille, cette fois ci incluant les effets d’une population de sous-halos dans le plan lentille et de halos distribués en ligne de visée. **Droite:** différence entre les deux simulations de lentille. Image de galaxie source provenant du recensement COSMOS (crédit: ESA/Hubble & NASA).

Dans ces deux éventualités, il s’agirait de l’une des plus importantes contraintes sur la nature de la particule de matière sombre à ce jour.

Cela étant dit, il est alors question d’avoir un moyen de détecter la signature de cette structure de matière sombre (ou son absence) dans des observations astronomiques. Puisque cette particule hypothétique n’interagit que par la force de gravité, la présence de halos de matière sombre peut être trahie par leurs effets gravitationnels. Notamment, la relativité générale prédit que la gravité dévie la trajectoire de la lumière; dans certaines circonstances, l’immense potentiel gravitationnel de halos massifs donne lieu au lentillage gravitationnel fort. Un exemple d’une modélisation de ce phénomène est illustré en figure 1.1, où l’image de gauche affiche la déviation extrême de la lumière d’une galaxie d’arrière-plan par la gravité d’une galaxie lentille d’avant-plan. En mesurant la géométrie de la déviation de la lumière de la source, il est possible d’inférer la masse et même la forme du halo de matière sombre de la lentille. Dans les dernières décennies, une série d’observations de systèmes lentillés ont été analysées afin de détecter avec confiance ces halos galactiques massifs.

De son côté, la signature de halos légers dans des systèmes lentillés est bien plus subtile, mais toutefois détectable. Par la sensibilité des images d’arcs étendus aux perturbations introduites par les halos légers, le lentillage gravitationnel fort représente également l’une des meilleures sondes de la distribution de structure sombre aux petites échelles. Les populations de structure d’un système de lentille peuvent être séparées en deux catégories: les sous-halos, se trouvant imbriqués dans l’étendue du grand halo galactique, ainsi que les halos de champ, indépendants de la galaxie lentille et distribués en ligne de visée. Les effets de ces deux populations dans une lentille gravitationnelle



**Fig. 1.2.** Schéma d’une population de halos légers dans un certain système de lentille gravitationnelle. La source se situe du côté gauche, émettant des rayons de lumière représentés par les traits rouges. Les trajectoires de ceux-ci sont principalement déviées par la galaxie lentille se trouvant dans le disque central, mais aussi par une multitude d’autres halos, avant de converger vers l’observateur du côté droit. Des halos de champ, distribués le long de la direction de décalage vers le rouge (*redshift*), sont représentés par des cercles, comme le sont les sous-halos dans le plan de la lentille. La taille et la couleur des cercles varie de façon proportionnelle à la masse et la concentration des halos, respectivement. Figure tirée de Gilman *et al.* (2022).

sont illustrés en l’image centrale de la figure 1.1, et un exemple de leur distribution spatiale typique est schématisé à la figure 1.2. Bien que cette simulation implique plus d’une dizaine de milliers de halos légers, la différence avec le modèle comprenant uniquement la galaxie lentille est à peine perceptible à l’œil nu.

L’exemple de la figure 1.1 met en évidence la difficulté de détecter la structure aux petites échelles dans les lentilles gravitationnelles. Une multitude de différentes populations de halos légers peuvent mener à des observations semblables, et il faut surtout parvenir à la déceler parmi les configurations possibles de lentille et de source. Puisque l’effet de ces halos sur l’image observée est si faible, l’analyse de la structure dans une seule lentille nécessite une énorme quantité de simulations. Hors, un très grand échantillon statistique de lentilles est requis pour contraindre des modèles de matière sombre de cette façon. Qui plus est, il a été projeté que des expériences observationnelles à venir telles que le télescope spatial Euclid (Refregier *et al.*, 2010; Laureijs *et al.*,

2011) et le recensement LSST de l’Observatoire Rubin (LSST Dark Energy Science Collaboration (LSST DESC) *et al.*, 2021) vont découvrir des centaines de milliers de nouvelles lentilles gravitationnelles (Collett, 2015). Dans ce régime de volume de données, l’étude de la structure sombre aux petites échelles dans les lentilles gravitationnelles devient possible, mais présente un imposant défi d’analyse statistique.

De tels problèmes statistiques se cadrent généralement dans le contexte d’inférence Bayésienne, où l’on cherche à obtenir la distribution à posteriori sur des paramètres d’intérêt conditionnée sur un ensemble d’observations. Les méthodes Bayésiennes exigent de pouvoir modéliser les données avec des simulateurs détaillés, mais aussi exacts. Dans les dernières années, l’étude de la structure aux petites échelles s’est principalement concentrée sur la détection de sous-halos dans le plan de la galaxie lentille, qui peuvent facilement être modélisés grâce au lentillage à plan unique. Dans le cas de la simulation du lentillage par de la structure en ligne de visée, on doit généralement faire appel au formalisme du lentillage gravitationnel à plans multiples. Cette approche, qui a servi à modéliser la seconde lentille de la figure 1.1, opère en traçant les rayons de lumière à travers l’entièreté du volume dans lequel sont distribués les halos de champs. On peut s’imaginer de trancher le volume en figure 1.2 le long de l’axe horizontal en plusieurs plans de lentille, et de tracer les rayons de lumière à travers une tranche de décalage vers le rouge à la fois. Cela prodigue une modélisation détaillée du lentillage par les halos de champ, au prix d’une dispendieuse opération récursive sur une multitude de plans de lentille. Cette nature récursive est une importante entrave aux problèmes d’inférence, qui nécessitent des centaines de milliers de simulations, et explique l’absence relative de littérature sur la détection de structure en ligne de visée dans les lentilles gravitationnelles.

Motivés en partie par cette difficulté, plusieurs auteurs ont développé différentes méthodes pour modéliser la contribution de la structure en ligne de visée dans les lentilles gravitationnelles. Certains approximent les inhomogénéités en ligne de visée avec un plan unique, les traitant comme des perturbations par effets de marée (Kovner, 1987; Bar-Kana, 1996; Schneider, 1997; Birrer *et al.*, 2017a) sous la forme d’une convergence externe et d’un cisaillement supplémenté à la galaxie lentille principale, ou bien en concevant un stratagème permettant de projeter les halos de champ dans le plan de lentille principal en adaptant leurs paramètres (Şengül *et al.*, 2020). D’autres ouvrages ont introduit des formalismes hybrides, rejoignant des aspects du lentillage à plans multiples et des perturbations par effets de marée (McCully *et al.*, 2014; Schneider, 2014; Fleury *et al.*, 2021a,b). Hélas, pour ce que ces formalismes approximatifs offrent en efficacité, ils souffrent en leur exactitude. Tel que remarqué par McCully *et al.* (2017), le lentillage à plans multiples introduit des effets non-linéaires qui ne peuvent pas être reproduits par les approximations à plan unique. Pour leur part,



les approches hybrides demeurent moins exactes que le véritable lentillage à plans multiples, tout en maintenant une certaine forme de récursion.

Dans ce mémoire, je présente une méthodologie alternative pour la simulation des effets de halos en ligne de visée dans les lentilles gravitationnelles. Ce nouveau formalisme permet de modéliser ces effets à un niveau de détail comparable au véritable lentillage gravitationnel à plans multiples, tout en étant accéléré par des modèles d'apprentissage profond. L'efficacité et l'exactitude de ce simulateur lui confèrent la viabilité requise pour les difficiles problèmes d'inférence ciblant la structure sombre aux petites échelles qui, grâce à l'énorme volume d'observations attendu dans la prochaine décennie, ont le potentiel de révéler les propriétés fondamentales de la matière sombre.

## 1.1. Description du mémoire

L'objectif principal de ce mémoire est de développer un formalisme permettant d'accélérer la simulation du lentillage gravitationnel à plans multiples grâce à des modèles d'apprentissage profond, dans le but de rendre plausibles les efforts d'inférence à venir concernant la structure de matière sombre aux petites échelles.

Le chapitre 2 est une introduction à la matière sombre, avec une perspective historique et un bref aperçu du problème contemporain de la structure aux petites échelles. Le chapitre 3 est une introduction à la théorie du lentillage gravitationnel fort et à son formalisme à plans multiples. Le chapitre 4 présente les fondements de l'apprentissage automatique, ainsi qu'une classe particulière de modèles du nom de réseaux autoattentifs. Enfin, le chapitre 5 présente en détails la méthodologie développée dans le cadre de ce mémoire, ainsi que des comparaisons de ses résultats à ceux du véritable lentillage à plans multiples.

## 1.2. Déclaration de l'étudiant

Je, Charles Wilson, déclare que l'entièreté du travail présenté dans ce mémoire est le mien. J'ai effectué la revue de littérature de ce mémoire et rédigé les chapitres introductifs 1, 2, 3 et 4. Les figures provenant de sources externes sont clairement identifiées.

Pour l'article présenté au chapitre 5, j'ai mené le développement d'une nouvelle architecture de réseau de neurones avec l'aide de Yashar Hezaveh, Adam Coogan, Daniel Gilman et Kyunghyun Cho. Ma contribution à ce projet est l'implémentation de cette architecture originale, la simulation de populations de halos et d'angles de déflexion effectifs, la production d'ensembles d'entraînement pour des modèles d'avant-plan et d'arrière-plan, le développement du code d'entraînement pour ces

modèles, leur entraînement même ainsi que la recherche d'hyper-paramètres et d'architectures, la production des résultats et finalement la rédaction de l'article.

## Chapitre 2

---

# Matière sombre

La matière sombre ça existe pas

---

ALICE LUSCHER

S'agirait-il simplement de la poursuite d'une chimère, de l'éther de notre époque? Comme il en a été le cas pour tant d'autres frontières passées de notre compréhension de l'Univers, la matière sombre n'est aujourd'hui qu'une théorie parmi plusieurs autres cherchant à expliquer les mesures de l'astrophysique contemporaine. Certaines propositions alternatives à cette particule inconnue viennent modifier la loi universelle de la gravitation de Newton, telles que MOND et sa généralisation relativiste TeVeS. Cependant, aucune de ces théories n'arrive à expliquer une aussi vaste gamme d'observations astronomiques que celle de la matière sombre. Cette particule hypothétique est l'un des ingrédients principaux du modèle standard de la cosmologie,  $\Lambda$ CDM. L'étude de sa nature fondamentale demeure l'un des efforts principaux de la physique moderne, avec une multitude d'expériences cherchant à détecter sa signature dans des détecteurs souterrains sur la Terre comme aux échelles extragalactiques. Ce chapitre présente un bref aperçu de l'histoire de la matière sombre en section 2.1, ainsi qu'une introduction au problème des petites échelles dans le modèle cosmologique actuel en section 2.2.

### 2.1. Historique

Les premières stipulations de l'existence d'une certaine *matière sombre* au sein de la communauté scientifique datent de plus d'un siècle, bien que ce à quoi l'on fait référence par le terme ait considérablement évolué au fil du temps. Dans une lettre portant sur les mouvements propres des étoiles Procyon et Sirius, Bessel (1844) commente que la lumière n'est pas une propriété intrinsèque à la

matière, et que l'existence d'innombrables étoiles visibles n'interdit pas en soi l'existence d'innombrables étoiles invisibles. Les astronomes Secchi (1877) et Barnard et Ranyard (1894) ont proposé au 19<sup>e</sup> siècle l'existence de nuages ou structures sombres distribués en l'espace, qui masqueraient des champs d'étoiles lumineux d'arrière-plan. Avec les premières analyses de dynamique stellaire dans la Voie Lactée par Lord Kelvin (1904), Öpik (1915) et Kapteyn (1922) au début du 20<sup>e</sup> siècle, le terme *matière sombre* en est venu à faire référence à toutes formes de nuages de gaz froid ou d'étoiles peu brillantes dans la galaxie. Leur masse s'ajouterait donc à celle des étoiles visibles pour expliquer les observations de dispersion des vitesses stellaires.

Ce sont les travaux révolutionnaires de Fritz Zwicky dans les années 1930 qui sonnèrent le glas de cette définition archaïque de la matière sombre. Dans un article étudiant les amas de nébuleuses extragalactiques (galaxies d'autrefois), Zwicky (1937) énumère plusieurs méthodes novatrices permettant l'estimation de leurs masses, en opposition aux estimés de l'époque principalement basés sur la comparaison de leur luminosité à  $L_{\odot}$ , celle du Soleil, et leur courbes de rotation képlériennes. Notamment, il emploie le théorème du viriel dans une analyse de l'amas de Coma pour obtenir une masse moyenne minimale par galaxie membre de l'amas d'au moins  $4.5 \times 10^{10} M_{\odot}$ . Prenant pour luminosité moyenne de galaxie d'amas une valeur de  $8.5 \times 10^7 L_{\odot}$ , cela indiquerait un immense rapport masse/luminosité d'environ  $\Upsilon = 500$ . Cette valeur<sup>1</sup> est difficile à avaler pour les contemporains de Zwicky; un aussi grand rapport masse/luminosité serait une indication que la majorité de la masse des galaxies ne serait pas contenue en sa matière brillante, mais devrait plutôt être attribuée à une composante sombre<sup>2</sup>, répandue entre les corps lumineux.

Dans ce même article de 1937, Zwicky propose une seconde méthode d'estimation de masse reposant sur une prédiction qui s'avérera à être remarquablement presciente. D'après des calculs publiés par Einstein l'année précédente (Einstein, 1936), la théorie de la relativité générale prédit la possibilité de formation d'images multiples d'une étoile par un effet de *lentillage gravitationnel* d'une étoile d'avant-plan. La quantité de déflexion de la lumière d'arrière plan (un angle  $\theta_E$  dans le plan du ciel, qui vint à prendre le nom de rayon d'Einstein) serait directement proportionnelle à la masse de l'étoile lentille:

$$\theta_E = \sqrt{\frac{4GM}{c^2 D}} \simeq 3 \left( \frac{M}{M_{\odot}} \right)^{1/2} \left( \frac{D}{\text{Gpc}} \right)^{-1/2} \mu\text{as}, \quad D = \frac{D_s D_{\ell}}{D_{\ell s}} \quad (2.1.1)$$

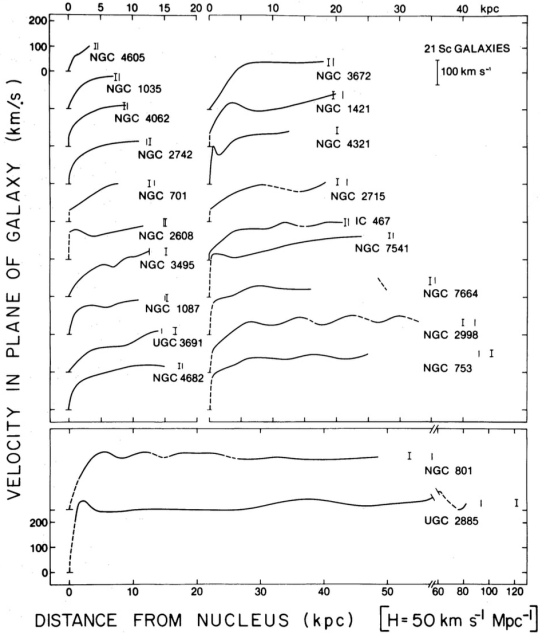
<sup>1</sup>Malgré que cet estimé soit fondé sur une valeur primitive de  $H_0$ , la constante de Hubble, une analyse avec une valeur de  $H_0$  plus actuelle mène à un  $\Upsilon$  d'environ 60, qui est toujours plus grand qu'attendu pour la matière lumineuse observée.

<sup>2</sup>Zwicky utilise toutefois encore *matière sombre* dans cet article pour référer à des étoiles froides, des corps solides micro- et macroscopiques ainsi que du gaz.

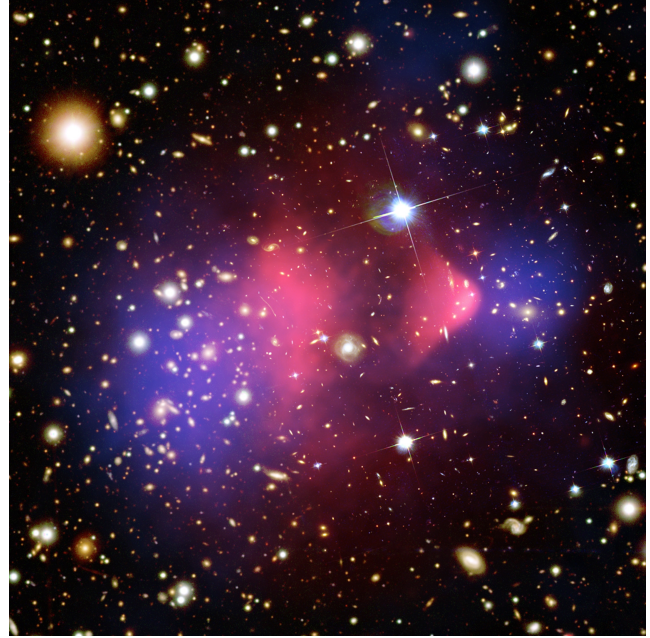
Einstein considèrerait trè̀s improbable l'observation de cet effet, les ètoiles ètant trop peu massives pour produire des dèflexions pouvant ètre rèsolues par les instruments de l'èpoque. Le point de vue de Zwicky ètait bien plus optimiste, soutenant que l'observation de lentilles gravitationnelles impliquant des galaxies, corps cèlestes beaucoup plus massifs que des ètoiles, ètait une certitude. De plus, cet effet de la thèorie d'Einstein prodiguerait une mètode de mesure directe et prècise des masses de galaxies. Cela s'est avèrè prophètique quarante ans plus tard, avec la dècouverte de la premièrè lentille gravitationnelle (Walsh *et al.*, 1979): le Quasar Jumeau QSO 0957 + 561, à un dècalage vers le rouge de  $z = 1.41$ , doublement lentillè à une sèparation angulaire de 6" par une galaxie à  $z = 0.355$ . Suite à 1979, le phènòmène de lentillage gravitationnel a ètè cimentè comme un formidable outil naturel ayant permis un nouvel essor de dècouvertes en astrophysique extragalactique, et qui est aujourd'hui central à l'ètude de la matièrè sombre. Le lentillage gravitationnel est traitè en de plus amples dètails au chapitre 3.

Reculant momentanément d'une dècennie, les annèes 1960 ont vu s'accumuler des observations èliminant le gaz comme explication de la "masse manquante" dans les amas de galaxies (Penzias, 1961; Meekins *et al.*, 1971). Si l'on porte en graphique la relation entre la vitesse orbitale de la matièrè visible d'une galaxie et sa distance radiale du centre de cette galaxie, la mècanique newtonienne prèdit que l'on devrait observer la vitesse dècroître avec la distance. Bien entendu, cela suppose que la majoritè de la masse d'une galaxie se trouve sous la forme d'ètoiles et de gaz dans son bulbe central. Les travaux de Vera Rubin ètudiant un èchantillon de galaxies spirales, effectués à l'observatoire de Kitt Peak au fil des annèes 1970, menèrent à la dècouverte que ces courbes de rotation ètaient surprenamment plates (Rubin *et al.*, 1978), une mesure rèvolutionnaire dont on peut voir des exemples en figure 2.1. Ces observations reprèsentent l'une des premièrès indications convaincantes que la plupart des galaxies seraient imbriquées dans de grands halos de matièrè invisible, tel l'avait proposè Zwicky en 1933. Afin d'expliquer la platitude systématique des courbes de rotation galactiques, ces halos devraient s'ètendre bien au-delà du rayon galactique et leur masse devrait largement dèpasser celle de la matièrè lumineuse. Tout ce remue-ménage fait sortir les thèoriciens de leurs tanières et donne naissance à leur nouvelle activité de choix: proposer des candidats expliquant la nature de la matièrè sombre. Une foulée d'objets astronomiques potentiellement responsables sont èventuellement réfutés à leur tour, tout semblant indiquer une nature non-baryonique à la matièrè sombre.

Parallèlement à ces dèveloppements a lieu la montèe en prominence de la cosmologie physique, propulsée par la dècouverte fortuite du fond diffus cosmologique (CMB, de l'anglais *cosmic microwave background*) par Penzias et Wilson (1965). Dans ce qui est aussi la renaissance de la



**Fig. 2.1.** Vitesses orbitales moyennes dans le plan galactique en fonction de la distance radiale du noyau pour un échantillon de 21 galaxies spirales de type Sc, arrangées en ordre croissant de rayon galactique. On y remarque que les courbes de vitesse deviennent "plates" au-delà de la région du bulbe central. Figure tirée de Rubin *et al.* (1980).



**Fig. 2.2.** Image composite de l'amas de la Balle, montrant le résultat d'une collision frontale entre deux amas de galaxies. Le gaz chaud de matière baryonique des amas est représenté en rose, imagé en rayons X par le télescope Chandra. Les halos de matière sombre de chaque amas, détectés par le lentillage gravitationnel faible d'objets d'arrière-plan, sont représentés en bleu. L'image dans le visible des galaxies de champ et membres de l'amas provient des télescopes Magellan et Hubble. Crédit: rayons X: NASA/CXC/CfA/Markevitch *et al.* (2002); lentillage: NASA/STScI, ESO WFI, Magellan/U.Arizona/Clowe *et al.* (2006); visible: NASA/STScI, Magellan/U.Arizona/Clowe *et al.* (2006)

relativité générale, plusieurs efforts sont fait pour estimer la fraction prise par la matière dans la densité d'énergie totale de l'Univers (Ostriker *et al.*, 1974; Einasto *et al.*, 1974), et la physique des particules s'en mêle en cherchant dans le Modèle Standard ou ailleurs des particules candidates de matière sombre. Une variété éclectique de tels candidats existe aujourd'hui dans la littérature, avec des propriétés et échelles énergétiques vastement différentes, mais partageant le trait commun de ne pas interagir avec la force électromagnétique (ou très peu). C'est à cette particule inconnue qui n'interagit pratiquement que par la gravité<sup>3</sup> que l'on attribue le sens moderne du terme *matière sombre*.

<sup>3</sup>Certaines particules candidates de matière sombre nommées WIMPs (*Weakly Interacting Massive Particle*) pourraient également interagir via la force nucléaire faible.

L'une des preuves empiriques les plus probantes à ce jour en appui de l'existence de la matière sombre est parvenue en 2004, avec l'analyse de la distribution de masse dans l'amas de la Balle (Clowe *et al.*, 2004). Comme il se voit en figure 2.2, les vestiges d'une titanesque collision entre deux amas de galaxies révèlent que leurs halos de matière sombre non-collisionnelle se seraient mutuellement traversés, alors que les nuages de gaz collisionnel du milieu intra-amas se seraient percutés de plein fouet pour être ralentis et séparés de leurs halos respectifs. En l'absence de halos de matière sombre, la majorité des effets de lentillage gravitationnel de cet amas seraient attribuables à ces nuages de gaz chaud, alors que les mesures indiquent hors de tout doute l'existence des énormes halos de chaque côté de ce système. Les théories alternatives de gravité modifiée telles que MOND et TeVeS, ayant été développées pour expliquer les courbes de rotation plates, ont beaucoup de difficulté à expliquer le décalage entre les deux maxima de lentillage faible et ceux de lumière en rayons X dans l'amas de la Balle. Supposant l'absence des halos de matière sombre dans l'amas, l'alignement fortuit de structures en ligne de visée pourrait expliquer ce phénomène, malgré que cette interprétation ne soit pas favorisée. Angus *et al.* (2006) parvient à réconcilier les observations de l'amas de la Balle avec MOND en invoquant la présence de matière non-collisionnelle sous la forme de neutrinos stériles<sup>4</sup>. Avec MOND ayant été initialement théorisée pour remédier au problème des courbes de rotations aux échelles galactiques, l'hypothèse d'une supplémentation par ces neutrinos stériles vient soulager les importantes tensions de MOND aux grandes échelles des amas de galaxies. Or, une publication récente (Banik *et al.*, 2023) étudiant les orbites de systèmes d'étoiles binaires larges dans les données galactiques de Gaia DR3 (Gaia Collaboration *et al.*, 2023) favorise à  $19\sigma$  des dynamiques Newtoniennes plutôt que MOND aux petites échelles sous-galactiques à l'étude, c.-à-d. de l'ordre de 10 kAU.

Ainsi, le débat cosmologique entre la matière sombre et MOND est loin d'être terminé, et une discussion détaillée de son état actuel dépasse le cadre de cet ouvrage. Chacune des théories connaissent un certain succès à expliquer des observations astrophysiques, mais à des échelles de distance vastement différentes. Il demeure toutefois que l'opinion scientifique actuel est majoritairement en faveur de la proposition de la matière sombre. Cependant, notre compréhension des systèmes astrophysiques étudiés pour mettre à l'épreuve ces théories demeure limitée en plusieurs aspects, par exemple la dynamique des galaxies. Les théories de la matière sombre et de MOND

---

<sup>4</sup>Les neutrinos stériles, tout comme la plupart des particules candidates de matière sombre, sont des particules hypothétiques n'ayant pas été détectées à ce jour. Ce type de neutrino est dit *stérile* car il n'interagirait qu'avec la gravité, contrairement aux neutrinos connus que l'on dit *actifs*, qui interagissent également avec la force faible.

sont toutes deux incomplètes et possèdent plusieurs degrés de liberté pour s’ajuster à des observations. À l’aube de cette nouvelle génération d’observatoires, seul l’avenir dira quelle théorie survivra aux découvertes à venir, et sous quelle forme.

## 2.2. La cosmologie moderne en crise

Suite à la découverte de l’expansion accélérée de l’Univers par Riess *et al.* (1998), le portrait généralement accepté de l’Univers se cadre dans le paradigme de  $\Lambda$ CDM (Carroll *et al.*, 1992). Ce modèle décrit un Univers doté d’une constante cosmologique positive  $\Lambda$  associée à une énergie sombre gouvernant l’accélération cosmique, de matière sombre froide (*Cold Dark Matter*, référée par CDM dans le reste du texte) et de matière baryonique ordinaire. Selon les résultats de 2018 de la collaboration du satellite Planck (Planck Collaboration *et al.*, 2020), la matière sombre froide compte pour 26% du bilan énergétique total de l’Univers, alors que l’on en attribue 5% à la matière baryonique et 68% à l’énergie sombre.  $\Lambda$ CDM a été célébré pour la confirmation observationnelle d’une multitude de ses prédictions, telles que la distribution de la structure aux grandes échelles de l’Univers, le spectre de puissance du CMB et la formation des galaxies. Cependant, de nombreuses remises en question du modèle persistent à ce jour, particulièrement que nous ne savons pas expliquer la nature de la constante cosmologique  $\Lambda$  ni de la CDM. D’autre part, il existe également une multitude de tensions observationnelles avec les prédictions de  $\Lambda$ CDM, qui menacent de remplacer ce paradigme par un modèle alternatif. En particulier, il sera ici question de la crise aux petites échelles.

J’identifie deux classes principales de modèles de matière sombre particulière, soit celle de la CDM, matière sombre froide, et de la WDM, matière sombre tiède (de l’anglais *Warm Dark Matter*). Ces modèles se distinguent principalement par la formation de structure dans le jeune Univers à travers les effets de la quantité  $\lambda_{\text{FS}}$ , l’échelle de dérive libre (*free-streaming length*). Cette échelle est définie comme la distance comobile caractéristique qu’une particule de matière sombre aurait pu parcourir depuis son découplage du plasma primordial<sup>5</sup> jusqu’à l’époque d’égalité matière-radiation  $t_{\text{eq}}$ , environ 50 000 ans plus tard, sans être significativement affectée par des interactions gravitationnelles (Schneider, 2012):

$$\lambda_{\text{FS}} = \int_0^{t_{\text{eq}}} \frac{\sigma(t)}{a(t)} dt \quad (2.2.1)$$

où  $a(t)$  est le facteur d’échelle caractérisant l’expansion de l’Univers, et  $\sigma(t)$  est la dispersion de vitesse d’une particule de matière sombre. C’est à cette époque que la structure de matière sombre

---

<sup>5</sup>qui s’approche à  $t = 0$ , étant quelques instants après le Big Bang.



croît par instabilité de Jeans, s’effondrant gravitationnellement pour former des halos. Cependant, l’échelle  $\lambda_{\text{FS}}$  est associée à la masse de Jeans à  $t_{\text{eq}}$ :

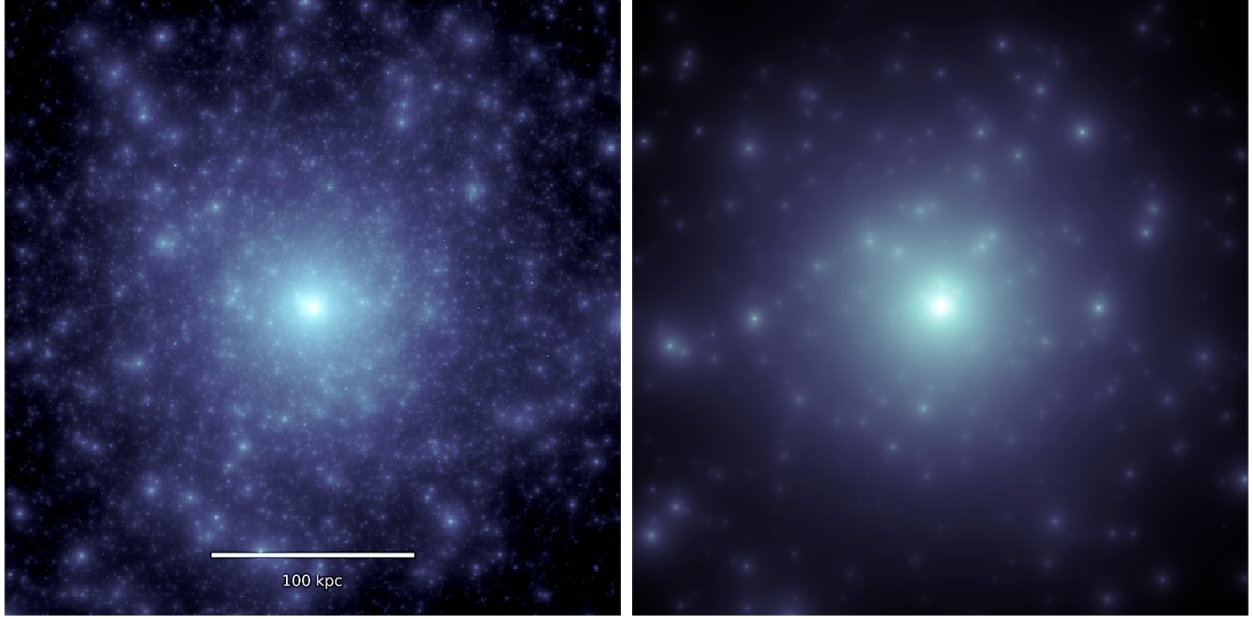
$$M_J = \frac{4\pi}{3} \rho_m \left( \frac{\lambda_{\text{FS}}}{2} \right)^3 \quad (2.2.2)$$

où  $\rho_m$  est la densité de matière sombre à  $t_{\text{eq}}$ . Au-deçà de l’échelle  $\lambda_{\text{FS}}$ , les perturbations de densité sont stables selon le critère de Jeans. La dispersion de vitesse a ainsi pour effet d’amortir ou de supprimer entièrement la croissance de structures. Cela entraîne une troncation au spectre de puissance primordial de la structure, indiquant une masse minimale  $M_J$  à laquelle des halos de matière sombre auraient pu être formés.

Dans les théories de type CDM, les particules candidates ont généralement des masses de l’ordre du GeV<sup>6</sup>. Ces masses importantes correspondent à des dispersions de vitesse  $\sigma$  (et conséquemment  $\lambda_{\text{FS}}$ , par (2.2.1)) négligeables. Du côté de WDM, il est sujet de particules de l’ordre du keV, maintenant des vitesses thermiques au-delà de l’époque dominée par la radiation et ayant des dispersions  $\sigma$  petites mais non-négligeables, d’où la description de matière sombre *tiède*. En somme, ces arguments indiquent des effets négligeables de  $\lambda_{\text{FS}}$  pour les théories CDM, alors qu’ils sont significatifs pour celles de type WDM.

L’abondance de halos de matière sombre dans un volume comobile en fonction de leur masse est décrite par la fonction de masse, et c’est principalement en cette fonction que les modèles CDM et WDM divergent. La fonction de masse de CDM démontre une certaine invariance d’échelle, alors que l’effacement de fluctuations aux petites échelles du spectre primordial par  $\lambda_{\text{FS}}$  a pour effet d’introduire un point tournant dans la fonction de masse de WDM, supprimant la puissance à partir des échelles de masse correspondant aux halos de galaxies satellites (Lovell *et al.*, 2014). Nombre de simulations cosmologiques à  $N$  corps de modèles CDM comme WDM ont démontrées que les halos massifs sont hôtes d’une certaine quantité de sous-halos (Ghigna *et al.*, 1998; Lovell *et al.*, 2014), des halos de plus faible masse ayant survécu au processus de fusion hiérarchique lors de la formation de structure. Ces sous-halos sont distribués selon leur propre fonction de masse dépendante de la masse du halo hôte, eux-mêmes distribués selon une fonction de masse globale. Toutefois, les effets de  $\lambda_{\text{FS}}$  affectent de façon semblable ces deux types de fonctions de masse, entraînant un écart important entre les abondances de sous-halos de faible masse entre CDM et WDM. Cette différence est illustrée dans la figure 2.3. J’utiliserai à partir d’ici le terme "halos

<sup>6</sup>C’est le cas des candidats de type WIMP comme le neutralino. Or, un autre candidat pour CDM est l’axion, une particule possiblement ultra-légère (les principales expériences étudient des axions de masses allant de  $10^{-22}$  eV/ $c^2$  à 1 eV/ $c^2$ ). Cette particule hypothétique pourrait perdre assez d’énergie cinétique lors d’une époque inflationnaire pour former un condensat de Bose-Einstein, lui conférant les propriétés de la matière sombre froide malgré sa petite masse.



**Fig. 2.3.** **Gauche:** réalisation de sous-structure dans un halo massif comparable à celui de la Voie Lactée ( $\sim 10^{12} M_{\odot}$ ), selon une fonction de masse correspondant à CDM. **Droite:** réalisation de sous-structure dans le même halo, cette fois selon une fonction de masse de WDM ayant un point tournant de troncation vers  $10^8 M_{\odot}$ . Crédit: Gilman (2020)

légers" pour faire référence aux halos de matière sombre de masse inférieure à  $10^{10} M_{\odot}$ , que ceux-ci soient des sous-halo d'un halo hôte plus massif ou bien des halos de champ indépendants.

Par sa dépendance à  $\lambda_{\text{FS}}$ , contraindre la fonction de masse de halos est l'une des approches les plus prometteuses pour contraindre les modèles particuliers de matière sombre. Pour ce faire, il est toutefois nécessaire d'avoir la capacité de détecter et mesurer les masses de ces halos légers. L'approche standard de correspondance d'abondance (de l'anglais *abundance matching*) repose sur l'hypothèse des modèles de formation de galaxies postulant que les celles-ci se forment à l'intérieur de halos de matière sombre. On peut alors mettre en correspondance l'abondance prédite de halos de masse  $M_h$  à l'abondance observée de galaxies de luminosité  $L_{\star}$  par une relation luminosité/masse de halo (Vale et Ostriker, 2004), ou bien à l'abondance observée de galaxies de masse  $m_{\star}$  par une relation masse stellaire/masse de halo (Behroozi *et al.*, 2013). Cependant, il existe d'importantes tensions entre les prédictions théoriques de CDM et les observations vers les échelles des galaxies naines; on voit dans la figure 2.3 que des simulations de CDM prédisent l'existence de milliers de sous-halos à travers une multitude d'échelles de masse dans des halos comme celui de la Voie Lactée, alors que les recensements observationnels n'ont identifié qu'une cinquantaine de galaxies satellites orbitant à l'intérieur du rayon de viriel de notre galaxie (Drlica-Wagner *et al.*, 2015) - il s'agit là du problème des satellites manquants.

Supposant que la matière sombre soit véritablement décrite par CDM, cette lacune serait une indication que l'efficacité de formation d'étoiles  $\epsilon_*$  devrait varier de façon non-linéaire en fonction de  $M_h$ , et que ces halos de matière sombre seraient trop peu massifs pour entraîner la formation d'étoiles (Sawala *et al.*, 2016). Alternativement, cela pourrait aussi indiquer qu'il n'existe tout simplement pas de halos de matière sombre à de telles échelles de masse, favorisant un modèle de type WDM. Les difficultés inhérentes à l'observation de matière lumineuse comme proxy pour tracer la distribution de matière sombre aux échelles sous-galactiques, telles que les effets incertains de perturbations par forces de marée et de rétroaction baryonique sur la structure de sous-halos (Despali et Vegetti, 2017), sont des importantes faiblesses de ce genre de méthodes dans l'effort de contraindre la nature de la matière sombre. Il y a ainsi un besoin criant pour des méthodes alternatives d'étudier la matière sombre aux petites échelles, afin de répondre à plusieurs des problèmes principaux auxquels le modèle cosmologique  $\Lambda$ CDM fait face aujourd'hui.



## Chapitre 3

---

### Lentillage gravitationnel fort

Passing paths that climb halfway into the  
void

As we cross from side to side, we hear the  
total mass retain

---

Jon Anderson

Le lentillage gravitationnel fort représente une sonde indépendante et bien plus directe de la distribution de matière sombre aux petites échelles. Par son couplage à seule la force de gravité, ce phénomène est sensible à la détection de structure même si celle-ci est complètement dépourvue de baryons sous forme d'étoiles ou de gaz (Cyr-Racine *et al.*, 2016). Dans des systèmes de lentilles galaxies-galaxies, l'immense potentiel gravitationnel du halo de la galaxie d'avant-plan va dévier la lumière de la source d'arrière-plan, en produisant plusieurs images magnifiées et distordues en anneau ou en arcs, selon l'alignement du système. En raison de cet effet de magnification, les images lentillées couvrent une plus grande région des environnements riches en sous-structure que sont les halos galactiques. Cela rend les images multiples d'une source lentillée susceptibles à des petites perturbations gravitationnelles locales par la sous-structure dans la galaxie lentille, faisant des systèmes de lentilles gravitationnelles fortes des laboratoires de choix pour l'étude des halos légers.

La détection de sous-halos individuels est possible par leur effet perturbatif sur les images d'arcs étendus caractéristiques de ce phénomène (Vegetti *et al.*, 2010, 2012; Hezaveh *et al.*, 2016). Par exemple, la présence d'un sous-halo près d'une des images lentillées peut être trahie par une perturbation gravitationnelle n'apparaissant pas dans les contre-images de la source. D'autre part,

l'inférence de population des perturbations introduites par des sous-halos dans un échantillon de galaxies lentillées permet de placer des contraintes sur la fonction de masse de la sous-structure (Vegetti *et al.*, 2014), un effort crucial vers la réconciliation des mesures aux petites échelles avec les prédictions de  $\Lambda$ CDM. Ces approches permettent également l'étude des halos de champ distribués en ligne de visée de lentilles galaxie-galaxie, comme le démontre la première découverte d'un halo de champ dans une lentille gravitationnelle par Şengül *et al.* (2022). Néanmoins, ce champ d'étude repose sur la modélisation de lentilles par le formalisme du lentillage à plans multiples, qui est significativement plus complexe que celui du lentillage à plan unique.

Ce chapitre présente plus formellement le lentillage gravitationnel fort et les équations qui le gouvernent. En section 3.1, le principe de Fermat est employé afin de déterminer la trajectoire d'un photon traversant l'espace-temps à proximité d'un potentiel gravitationnel. Cela permet de dériver l'expression des angles de déflexion, la quantité principale derrière le formalisme du lentillage gravitationnel. En section 3.2, l'équation de lentille est dérivée pour décrire un système de lentille galaxie-galaxie à plan unique. Enfin, la section 3.3 présente le formalisme de lentillage gravitationnel à plans multiples, généralisant l'équation de lentille pour décrire plusieurs plans de déflexions consécutives.

### 3.1. Principe de Fermat et angles de déflexion

Dans un référentiel de Lorentz local typique, la longueur d'onde de la lumière est significativement inférieure aux échelles de distortion de l'espace-temps; cela implique que les effets associés à la nature ondulatoire de la lumière - tels que sa diffraction par des distortions d'espace-temps - peuvent être ignorés. Cela justifie un traitement du lentillage gravitationnel fort reposant sur l'approximation de l'optique géométrique. Dans cette approximation, les équations décrivant la déviation de la lumière par une distribution de masse peuvent être dérivées à l'aide du principe de Fermat. Cette dérivation est largement inspirée de celles des manuels de Petters *et al.* (2001) et Meneghetti (2021).

Supposons qu'un photon émit par une source distante en un point de départ  $A$  est dévié par un potentiel gravitationnel,  $\Phi$ , en se propageant dans l'espace-temps, avant d'atteindre un point d'arrivée  $B$ . De façon analogue à la géométrie optique, on introduit ici un indice de réfraction  $n$  associé à la lentille gravitationnelle. Le principe de Fermat stipule que le photon doit suivre une trajectoire  $\mathbf{q}(\lambda)$  telle que la durée du parcours soit stationnaire, où  $\lambda$  paramétrise la trajectoire. En calcul des variations, cela correspond à des extrema de la durée de parcours:

$$\delta T = \delta \int_A^B n(\mathbf{q}(\ell)) \frac{d\ell}{c} = 0, \quad (3.1.1)$$

où  $\ell$  est un élément de longueur sur la trajectoire. Il s'agit alors de trouver une forme appropriée pour l'indice de réfraction. À cette fin, on fait appel au formalisme de la relativité générale pour décrire la métrique de l'espace-temps à proximité de la lentille. Supposons un Univers décrit par une cosmologie de Friedmann-Lemaître-Robertson-Walker (FLRW) de courbure nulle. Une lentille gravitationnelle représente une perturbation locale à la métrique FLRW par un potentiel newtonien  $\Phi(\mathbf{q})$  considéré faible ( $|\Phi|/c^2 \ll 1$ ). La géométrie de l'espace temps près de la lentille est alors approximée par la métrique de "champ faible":

$$ds^2 = g_{\mu\nu} dq^\mu dq^\nu = - \left(1 + \frac{2\Phi}{c^2}\right) c^2 d\tau^2 + a_L^2 \left(1 - \frac{2\Phi}{c^2}\right) d\mathbf{q}^2 \quad (3.1.2)$$

où  $a_L(t(\tau_L))$  est le facteur d'échelle à la coordonnée de temps  $t_L$ , qui correspond au temps propre  $\tau_L$  où la trajectoire du photon traverse la lentille, et  $d\mathbf{q}^2$  est l'intervalle métrique euclidien en  $\mathbb{R}^3$ . Le principe de Fermat est un cas particulier du principe de moindre action pour les particules sans masse. Les trajectoires physiques de telles particules, des photons par exemple, sont des géodésiques nulles, correspondant à un intervalle d'espace-temps  $ds = 0$ :

$$\left(1 + \frac{2\Phi}{c^2}\right) c^2 d\tau^2 = a_L^2 \left(1 - \frac{2\Phi}{c^2}\right) d\mathbf{q}^2. \quad (3.1.3)$$

La célérité efficace de la lumière dans ce champ gravitationnel est donc

$$v = \left| a(\tau) \frac{d\mathbf{q}}{d\tau} \right| = c \sqrt{\frac{1 + \frac{2\Phi}{c^2}}{1 - \frac{2\Phi}{c^2}}} \approx c \left(1 + \frac{2\Phi}{c^2}\right) \quad (3.1.4)$$

où la limite de champ faible  $|\Phi|/c^2 \ll 1$  a été supposée. L'indice de réfraction est donc simplement

$$n = \frac{c}{v} = \frac{1}{1 + \frac{2\Phi}{c^2}} \approx 1 - \frac{2\Phi}{c^2}. \quad (3.1.5)$$

Réécrivant l'élément de longueur  $d\ell$  en fonction d'un élément de paramètre de trajectoire  $d\lambda$ , nous avons

$$d\ell = \left| \frac{d\mathbf{q}}{d\lambda} \right| d\lambda. \quad (3.1.6)$$

Revenant à la condition d'extrema (3.1.1),

$$0 = \delta \int_{\lambda_A}^{\lambda_B} n(\mathbf{q}) \left| \frac{d\mathbf{q}}{d\lambda} \right| d\lambda \quad (3.1.7)$$

On reconnaît ici la forme d'une variation de la fonctionnelle de l'action  $\delta S(\mathbf{q}) = \delta \int_{\lambda_A}^{\lambda_B} \mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}; \lambda) d\lambda$ , où le lagrangien du photon s'écrit  $\mathcal{L} = n(\mathbf{q}) \sqrt{\dot{\mathbf{q}}^2}$ . La trajectoire du photon qui minimise l'action,

satisfaisant (3.1.7), est également une solution de l'équation d'Euler-Lagrange:

$$\frac{d}{d\lambda} \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{q}}} - \frac{\partial \mathcal{L}}{\partial \mathbf{q}} = 0 \quad (3.1.8)$$

Pour le lagrangien d'intérêt, celle-ci se développe à

$$\frac{d}{d\lambda} \left( n \frac{\dot{\mathbf{q}}}{|\dot{\mathbf{q}}|} \right) - |\dot{\mathbf{q}}| \nabla n = 0 \quad (3.1.9)$$

La quantité  $\dot{\mathbf{q}}$  est un 3-vecteur tangent à la trajectoire du photon en tout point  $\lambda$ . Nous avons un degré de liberté de choisir la paramétrisation de la trajectoire de sorte à ce que  $|\dot{\mathbf{q}}| = 1$  en tout point de celle-ci. Reprenant (3.1.9),

$$\begin{aligned} \frac{d}{d\lambda} (\dot{\mathbf{q}} n) &= \nabla n \\ \ddot{\mathbf{q}} n &= \nabla n - (\nabla n \cdot \dot{\mathbf{q}}) \dot{\mathbf{q}} \\ \ddot{\mathbf{q}} n &= \nabla n - (\nabla n \cdot \mathbf{e}_{\parallel}) \mathbf{e}_{\parallel} \end{aligned} \quad (3.1.10)$$

où on a défini le vecteur unitaire  $\mathbf{e}_{\parallel} = \dot{\mathbf{q}}$  parallèle à la trajectoire.

Nous allons ici invoquer l'approximation de Born du premier ordre, qui suppose les trajectoires de photons dans un système de lentille gravitationnelle galaxie-galaxie comme des lignes droites entre les perturbations de masse (c.-à-d. des géodésiques nulles dans la métrique FLRW non-perturbée d'arrière plan) qui sont déviées instantanément lorsqu'elles traversent de telles perturbations. Cela correspond à **l'approximation de la lentille mince**, où les distributions de masse tridimensionnelles déviant les trajectoires de photons sont projetées dans un **plan de lentille** perpendiculaire à la ligne de visée depuis un observateur, tel qu'illustré en fig. 3.1. Cette approximation est justifiée puisque les distances cosmologiques de l'ordre du Gpc qui séparent une source d'arrière-plan d'un plan lentille, et un plan lentille d'un observateur, sont bien plus grandes que l'échelle d'étendue d'une distribution de masse le long de la ligne de visée, qui est plutôt de l'ordre du kpc. Cela permet d'approximer que la distribution de masse est entièrement contenue en un plan de lentille.

Suivant ce raisonnement, la déviation du photon par  $\Phi$ , c.-à-d. son accélération instantanée  $\ddot{\mathbf{q}}$ , est contenue dans le plan de directions perpendiculaires à celle tangente à sa trajectoire. Le second terme du membre de droite de l'équation (3.1.10) est la projection du gradient de  $n$  en direction  $\mathbf{e}_{\parallel}$ , qui est soustrait au gradient complet pour ne laisser que le gradient en direction perpendiculaire



$\nabla_{\perp}$ . En introduisant la forme de  $n$  trouvée en (3.1.5),

$$\begin{aligned}\ddot{\mathbf{q}}_{\perp} &= \frac{1}{n} \nabla_{\perp} n = \nabla_{\perp} \log n = \nabla_{\perp} \log \left( 1 - \frac{2\Phi}{c^2} \right) \\ \ddot{\mathbf{q}}_{\perp} &\approx -\frac{2}{c^2} \nabla_{\perp} \Phi\end{aligned}\tag{3.1.11}$$

L'angle de déflexion du photon s'écrit alors

$$\hat{\alpha} = - \int_{\lambda_A}^{\lambda_B} \ddot{\mathbf{q}}_{\perp} d\lambda = \frac{2}{c^2} \int_{\lambda_A}^{\lambda_B} \nabla_{\perp} \Phi d\lambda\tag{3.1.12}$$

Les angles de déflexion caractéristiques pour des lentilles à l'échelle de galaxies sont de l'ordre de la seconde d'arc, ou de  $1/3600^{\circ}$  dans le plan du ciel. Cela est cohérent avec la limite de champ faible, valide pour les petits angles de déflexion. Au moment du périapse du photon par rapport à la lentille, c.à-d. le point de sa trajectoire ou leur distance est minimale, le vecteur séparant le photon du centre de la distribution de masse sera toujours perpendiculaire à la direction instantanée du photon. En fait, ce vecteur sera contenu dans le plan défini par les directions perpendiculaires à la trajectoire du photon. Dans la limite de champ faible et de lentille mince que nous utilisons, ce plan peut être approximé à être le même que le plan de lentille. On peut ainsi intégrer (3.1.12) supposant une trajectoire non-perturbée, le long de la coordonnée comobile  $\zeta$  parallèle à la ligne de visée entre la source et la lentille:

$$\hat{\alpha} = \frac{2}{c^2} \int_{-\infty}^{+\infty} \nabla_{\perp} \Phi d\zeta\tag{3.1.13}$$

On construit maintenant un système de coordonnées comobiles  $\mathbf{q} = (x, y, \zeta)$ , avec  $(x, y)$  les deux axes de coordonnées sur le plan de lentille, ayant pour origine le point où l'axe de la ligne de visée  $\zeta$  traverse le centre de la distribution de masse. On introduit ici le paramètre d'impact  $\boldsymbol{\xi} = (x, y)$ , définit comme le vecteur séparant le photon de la lentille lorsqu'il traverse le plan (voir fig. 3.1). D'après l'équation de Friedmann pour notre métrique FLRW perturbée (3.1.2), la perturbation de masse de la lentille satisfait l'équation de Poisson

$$\nabla^2 \Phi = 4\pi G \rho.\tag{3.1.14}$$

Supposons que la lentille à l'étude soit une masse ponctuelle, avec une densité associée s'écrivant  $\rho(\mathbf{q}) = M\delta^3(\mathbf{q})$ , où  $\delta$  est la fonction delta de Dirac. Le potentiel gravitationnel généré par cette densité qui satisfait l'équation de Poisson (3.1.14) est  $\Phi(\mathbf{q}) = -\frac{GM}{|\mathbf{q}|} = -\frac{GM}{\sqrt{\xi^2 + \zeta^2}}$ . L'angle de

déflexion au paramètre d'impact  $\xi$  se trouve alors par (3.1.13):

$$\hat{\alpha}(\xi) = -\frac{2GM}{c^2} \int_{-\infty}^{+\infty} \frac{\partial}{\partial \xi} \frac{1}{\sqrt{\xi^2 + \zeta^2}} d\zeta$$

$$\boxed{\hat{\alpha}(\xi) = \frac{4GM}{c^2 \xi^2} \xi} \quad (3.1.15)$$

Il s'avère que le résultat de la relativité générale trouvé ci-haut excède d'un facteur 2 ce qui s'obtiendrait par une dérivation newtonienne.

Il est particulièrement intéressant de remarquer que par la linéarité en  $M$  de  $\hat{\alpha}$  dans (3.1.15), on peut superposer linéairement les déflexions d'une multitude de lentilles se trouvant dans le même plan lentille. De plus, cette propriété permet de généraliser (3.1.15) à des distributions de masses étendues. À cet effet, il est utile de définir la densité surfacique

$$\Sigma(\xi) = \int_{\mathbb{R}} \rho(\xi, \zeta) d\zeta \quad (3.1.16)$$

où  $\rho$  est une densité de masse tridimensionnelle qui se voit projetée dans le plan lentille. Dans l'approximation de lentille mince, l'angle de déflexion total en un point  $\xi$  est le résultat d'une somme sur tous les éléments de masse  $dM = \Sigma(\xi') d^2 \xi'$ , correspondant à la convolution

$$\hat{\alpha}(\xi) = \frac{4G}{c^2} \int_{\mathbb{R}^2} \frac{(\xi - \xi')}{|\xi - \xi'|^2} \Sigma(\xi') d^2 \xi' \quad (3.1.17)$$

## 3.2. Équation de lentille

Dans un système de lentille gravitationnelle ayant une source à un décalage vers le rouge  $z_s$ , une lentille à  $z_\ell$  et un observateur à  $z_o = 0$  ( $z_s > z_\ell > z_o$ ), l'angle de déflexion représente une transformation non-linéaire entre les coordonnées angulaires du plan de source ( $\beta$ ) et celles du plan de lentille ( $\theta$ ), aussi appelé plan image. Par la nature angulaire des coordonnées, elles peuvent être reliées à des distances physiques sur leurs plans respectifs par des distances de diamètre angulaire. Ainsi, le paramètre d'impact d'un photon sur le plan de lentille serait  $\xi = D_\ell \theta$ , où  $D_\ell$  est la distance de diamètre angulaire de  $z_o = 0$  à  $z_\ell$  dans la métrique d'arrière-plan FLRW de courbure nulle<sup>1</sup>:

$$D_\ell = \frac{c}{H_0(1+z_\ell)} \int_0^{z_\ell} \frac{dz'}{\sqrt{\Omega_{r,0}(1+z')^4 + \Omega_{m,0}(1+z')^3 + \Omega_{\Lambda,0}}} \quad (3.2.1)$$

<sup>1</sup>Les relations angle-distance de la forme  $\xi = D_\ell \theta$  demeurent valides dans un espace-temps courbe par des modifications appropriées à l'expression de distance de diamètre angulaire (3.2.1).

où la constante de Hubble  $H_0$  est le taux d'expansion actuel de l'Univers et les quantités  $\Omega_{r,0}$ ,  $\Omega_{m,0}$  et  $\Omega_{\Lambda,0}$  sont les paramètres de densité actuels de la radiation, la matière et l'énergie sombre, respectivement. Les valeurs des paramètres du modèle cosmologique  $\Lambda$ CDM telles que mesurées par la Planck Collaboration *et al.* (2020) sont présentées dans l'annexe A. Cette mesure de distance est également définie du point de vue de plans à  $z \neq 0$  en changeant de borne inférieure d'intégration, p. ex. pour la distance de diamètre angulaire entre le plan de lentille et le plan source:  $D_{\ell s} \mapsto \int_{z_\ell}^{z_s}$ .

Il est utile d'établir un formalisme de lentillage gravitationnel entièrement en coordonnées angulaires, en particulier pour la forme (3.1.17) des angles de déflexion. À partir d'ici, les relations seront exprimées en terme de coordonnées angulaires. À cet effet, on définit une densité surfacique adimensionnelle  $\kappa$  nommée convergence:

$$\kappa(\boldsymbol{\theta}) = \frac{\Sigma(D_\ell \boldsymbol{\theta})}{\Sigma_{\text{cr}}}, \quad \Sigma_{\text{cr}} = \frac{c^2}{4\pi G} \frac{D_s}{D_\ell D_{\ell s}} \quad (3.2.2)$$

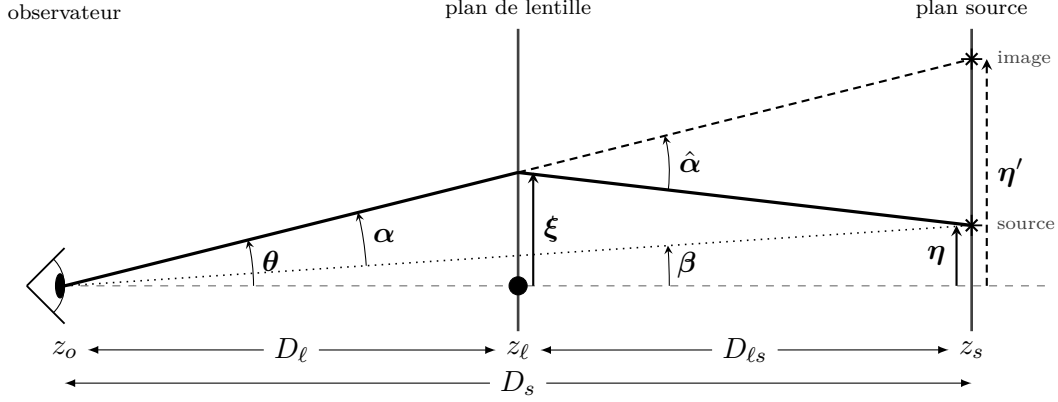
où  $\Sigma_{\text{cr}}$  est une densité surfacique critique, dépendant de la géométrie d'un système de lentille. Généralement, il peut être dit qu'avec un alignement suffisant entre source et lentille, la condition de supercriticalité  $\kappa > 1$  est suffisante pour mener à la formation d'images multiples de la source<sup>2</sup>. C'est par cette propriété que le lentillage gravitationnel fort se distingue du lentillage gravitationnel faible, qui ne fait que déformer une unique image de source sans la multiplier. La convergence permet de décrire entièrement la distribution de masse d'un plan de lentille, et conséquemment l'angle de déflexion total en tout point du plan:

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\theta}) = \frac{1}{\pi} \frac{D_s}{D_{\ell s}} \int_{\mathbb{R}^2} \kappa(\boldsymbol{\theta}') \frac{\boldsymbol{\theta} - \boldsymbol{\theta}'}{|\boldsymbol{\theta} - \boldsymbol{\theta}'|^2} d^2 \boldsymbol{\theta}' \quad (3.2.3)$$

Une source à une position angulaire véritable  $\boldsymbol{\beta}$  qui subit un effet de lentillage produira une image qui sera observée à la position  $\boldsymbol{\theta}$ . Dans l'approximation des petits angles, on peut se référer

---

<sup>2</sup>Pour des distributions de masse axisymétriques, la condition  $\kappa > 1$  est suffisante et nécessaire pour la formation de plusieurs images. Dans le cas de lentilles plus réalistes non-axisymétriques, cette condition est suffisante mais pas nécessaire, puisque des effets de cisaillement peuvent contribuer à la formation d'images multiples (Congdon et Keeton, 2018).



**Fig. 3.1.** Schéma d'une lentille gravitationnelle galaxie-galaxie dans l'approximation de la lentille mince. Le trait noir épais représente la trajectoire d'un photon qui se fait dévier par la lentille du plan en  $z_\ell$ .

au schéma en figure 3.1 pour relier ces positions par un simple argument de géométrie euclidienne:

$$\begin{aligned}
 D_s \beta &= \eta \\
 D_s \theta &= \eta' \\
 D_{\ell s} \hat{\alpha} &= \eta' - \eta \\
 D_s \beta &= D_s \theta - D_{\ell s} \hat{\alpha} \\
 \implies \beta &= \theta - \frac{D_{\ell s}}{D_s} \hat{\alpha}
 \end{aligned} \tag{3.2.4}$$

En définissant l'angle de déflexion réduit  $\alpha = \frac{D_{\ell s}}{D_s} \hat{\alpha}$  par rapport à l'observateur, on obtient **l'équation de lentille à plan unique**

$$\boxed{\beta = \theta - \alpha(\theta)} \tag{3.2.5}$$

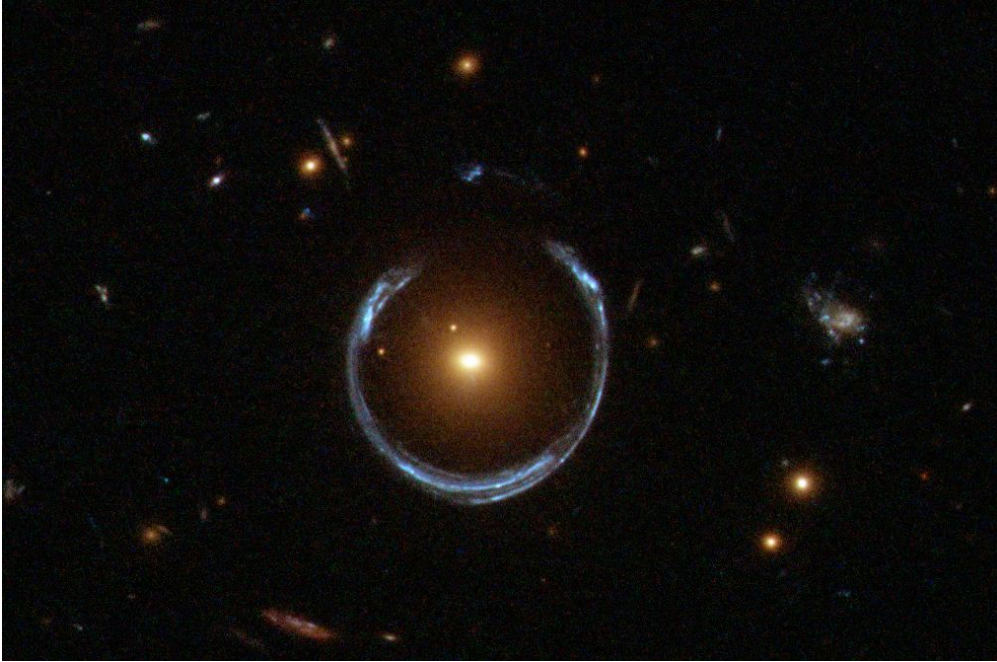
### 3.2.1. Traçage de rayons inverse

La relation (3.2.5) est l'équation maîtresse du lentillage gravitationnel fort, permettant la simulation d'images lentillées par le traçage de rayons inverse. Puisque  $\theta$  ne peut être mis en évidence dans (3.2.5) du point de vue de l'observateur, cet algorithme trace les rayons de lumière sur une grille pixellisée à partir de l'observateur jusqu'à la source, soit dans le sens inverse à la véritable trajectoire des photons. Tous les photons observés doivent obligatoirement converger à l'observateur, correspondant à un unique point dans l'espace. Ainsi, tracer les rayons dans leur sens véritable depuis la source à l'observateur impliquerait de pouvoir prédire quels chemins vont mener à l'observateur. Il est bien plus simple de plutôt tracer les rayons qui partent de l'observateur et frappent le plan de source, une cible pouvant être rendue arbitrairement grande.

En pratique, les rayons sont lancés depuis l’origine à  $z_o$  pour définir une grille de positions angulaires  $\theta$ , et sont ensuite fléchis par l’équation de lentille (3.2.5) pour obtenir une grille de positions angulaires distordues  $\beta$ . Un profil de lumière de la source peut ensuite être évalué sur ces coordonnées pour produire l’image lentillée.

### 3.3. Lentillage gravitationnel à plans multiples

Bien que la plupart des systèmes de lentilles gravitationnelles peuvent être modélisés adéquatement par un unique plan de lentille contenant une galaxie massive et possiblement des sous-halos, la réalité implique aussi la contribution de toute la structure distribuée en ligne de visée à la déviation entière de la lumière. Ces halos de champ légers sont isolés de halos galactiques et donc moins susceptibles aux effets de rétroaction baryonique (Duffy *et al.*, 2010). Ils devraient avoir un effet significatif sur les observables de lentillage (Chen *et al.*, 2003), pouvant même dominer le signal de sous-structure dans certains régimes (Li *et al.*, 2017). La caractérisation des effets de halos en ligne de visée dans les lentilles gravitationnelles, en conjonction avec ceux des sous-halos, permet



**Fig. 3.2.** SDSS J1148+1930, aussi appelée le fer à cheval cosmique, est un exemple impressionnant d’une lentille gravitationnelle galaxie-galaxie. La distante galaxie bleue à  $z_s = 2.381$  apparaît en plusieurs images fortement distordues, formant un anneau presque parfait d’un énorme rayon d’Einstein d’environ  $5''$ . La galaxie rouge à  $z_\ell = 0.444$  est responsable de cet effet de lentillage; avec une masse estimée de  $5.2 \times 10^{12} M_\odot$  (Schuldt *et al.*, 2019), il s’agirait de l’une des galaxies les plus massives jamais observées. Composite d’images en visible et infrarouge prises par l’instrument WFC3 de HST. Crédit: ESA/Hubble & NASA

de sonder les plus petites échelles massiques de la structure, ce qui en fait une des méthodes les plus probantes pour contraindre les modèles particulières de matière sombre.

Dans l'approximation de lentille mince, les sous-halos distribués en périphérie d'une lentille principale à un certain  $z$  sont considérés comme faisant partie du même plan de lentille que cette dernière. Par la propriété de superposition de masses démontrée avec l'équation (3.1.15), un tel système de lentille gravitationnelle peut être traité avec le formalisme du lentillage à plan unique. Hors, le lentillage par la structure en ligne de visée ne peut être traité de la même façon que celui par la structure dans le plan de la lentille. Dans un tel cas, les rayons de lumière sont déviés à une multitude de plans de lentilles distribués à différents  $z$ , ce qui nécessite un traitement mathématique plus méticuleux. Le formalisme du **lentillage gravitationnel à plans multiples** est une généralisation de celui à plan unique élaboré aux sections 3.1 et 3.2, où plusieurs plans de lentilles minces<sup>3</sup> parallèles parsèment l'axe de visée, chacun pouvant être hôte d'un nombre indéfini de halos. Le traitement qui suit est tiré de près de celui du manuel de Congdon et Keeton (2018).

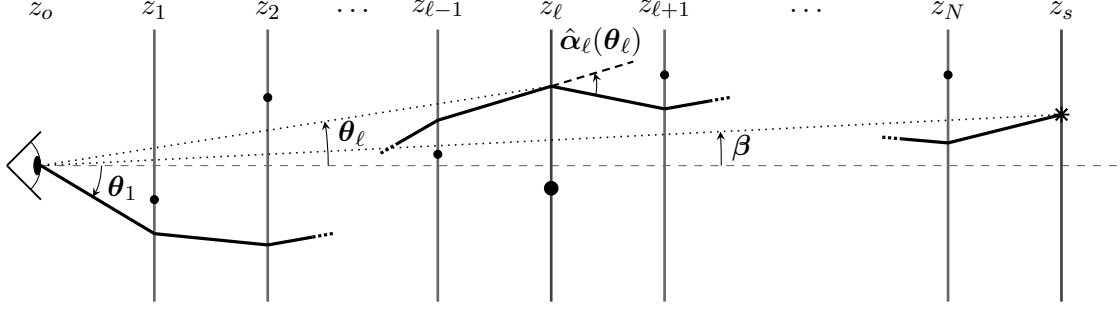
Considérons en premier lieu le lentillage à plans doubles, l'extension naturelle du lentillage à plan unique. Supposons deux plans de lentille aux décalages vers le rouge  $z_1, z_2$  ( $z_1 < z_2$ ) et distances de diamètre angulaire associées  $D_1, D_2$  par rapport à  $z_o$ . Invoquant le traçage de rayons inverses décrit en 3.2.1, un rayon de lumière tracé depuis l'observateur avec une direction initiale  $\boldsymbol{\theta}_1$  subira d'abord une déflexion  $\hat{\boldsymbol{\alpha}}_1(D_1\boldsymbol{\theta}_1)$ , suivie d'une déflexion  $\hat{\boldsymbol{\alpha}}_2(D_2\boldsymbol{\theta}_2)$ , avant d'atteindre le plan source en  $z_s$  à la position  $\boldsymbol{\beta}$ . Généralisant les arguments géométriques illustrés par la figure 3.1 et ayant mené à (3.2.4), on obtient la relation

$$\begin{aligned} D_s\boldsymbol{\beta} &= D_s\boldsymbol{\theta}_1 - D_{1s}\hat{\boldsymbol{\alpha}}_1(D_1\boldsymbol{\theta}_1) - D_{2s}\hat{\boldsymbol{\alpha}}_2(D_2\boldsymbol{\theta}_2) \\ \boldsymbol{\beta} &= \boldsymbol{\theta}_1 - \frac{D_{1s}}{D_s}\hat{\boldsymbol{\alpha}}_1(\boldsymbol{\theta}_1) - \frac{D_{2s}}{D_s}\hat{\boldsymbol{\alpha}}_2(\boldsymbol{\theta}_2) \\ \boldsymbol{\beta} &= \boldsymbol{\theta}_1 - \boldsymbol{\alpha}_1(\boldsymbol{\theta}_1) - \boldsymbol{\alpha}_2(\boldsymbol{\theta}_2) \end{aligned} \tag{3.3.1}$$

où on a invoqué la définition de l'angle de déflexion réduit, ici défini respectivement à chaque plan. Pour le lentillage à plans multiples, on définit le plan image comme celui au plus petit  $z$  avant l'observateur à  $z_o$ . Dans ce cas-ci, cela ferait de  $\boldsymbol{\theta}_1$  la coordonnée de l'image observée sur le plan image. De son côté, la coordonnée  $\boldsymbol{\theta}_2$  serait celle de l'image observée en l'absence du plan de lentille en  $z_1$ ; on peut alors imaginer  $\boldsymbol{\theta}_2$  comme étant la position de la source dans l'équation de lentille à

---

<sup>3</sup>L'approximation de lentille mince demeure valide dans ce scénario à condition que les distances inter-plans soient plus grandes que l'échelle des lentilles individuelles.



**Fig. 3.3.** Schéma du lentillage à plans multiples pour une lentille gravitationnelle galaxie-galaxie avec de la structure en ligne de visée. Le trait noir épais représente la trajectoire d'un photon.

plan unique (3.2.4) qui subirait une déflexion  $\hat{\alpha}_1(D_1\theta_1)$ :

$$D_2\theta_2 = D_2\theta_1 - D_{12}\hat{\alpha}_1(D_1\theta_1)$$

$$\theta_2 = \theta_1 - \frac{D_{12}}{D_2}\hat{\alpha}_1(\theta_1) \quad (3.3.2)$$

D'après cette relation entre les coordonnées  $\theta_2$  et  $\theta_1$ , il se voit que l'on peut retrouver en (3.3.1) l'équation de lentille à plan simple lorsque la distance entre les deux plans est nulle, e.g.  $D_{12} = 0$ , donc  $\theta_2 = \theta_1$ . Alors, les déflexions dans (3.3.1) se superposeraient linéairement à la même coordonnée angulaire.

Dans l'élaboration du formalisme de lentillage gravitationnel à plans multiples, nous nous intéressons au cas général de  $N$  plans lentille aux décalages vers le rouge  $z_i$ , avec  $i = 1, 2, \dots, N$  et  $z_i < z_j$  pour  $i < j$ . Le plan source à  $z_s$  est identifié par l'indice  $s \equiv N + 1$ , et l'observateur est toujours en  $z_o = 0$ . Ce scénario est illustré à la figure 5.1, où l'on a placé qu'une seule lentille par plan à des fins de présentation. La généralisation des équations à plans doubles (3.3.1) et (3.3.2) mène aux équations du lentillage gravitationnel à plans multiples:

$$\beta = \theta_1 - \sum_{i=1}^N \frac{D_{is}}{D_s} \hat{\alpha}_i(\theta_i) \quad (3.3.3)$$

$$\theta_i = \theta_1 - \sum_{m=1}^{i-1} \frac{D_{mi}}{D_i} \hat{\alpha}_m(\theta_m) \quad (3.3.4)$$

La relation récursive (3.3.4) fait en sorte que les coordonnées  $\theta_i$  à un certain  $z_i$  dépendent des déflexions par la structure à tous les plans précédents. On peut penser du phénomène décrit par l'équation (5.2.4) comme analogue à regarder à travers une loupe à travers la lentille d'une autre loupe, elle-même à travers d'une autre loupe, et ainsi de suite.

Cet aspect itératif présente une difficulté intrinsèque à ce formalisme, qui a des implications importantes pour l’algorithme de traçage de rayons inverse à plans multiples. Par exemple, afin de simuler les déflexions causées par une certaine population de halos en ligne de visée d’une lentille galaxie-galaxie, il serait nécessaire de tracer les rayons de lumière entre chaque plan de lentille en ordre croissant de  $z$ , un à la fois. De plus, le fait de changer ne serait-ce qu’un seul halo à un  $z_i$  près de  $z_o$  impliquerait de devoir recalculer toutes les trajectoires des rayons au-delà de ce  $z_i$ . Si l’on suppose une grille carrée de coordonnées angulaires de  $n^2$  pixels, en plus de choisir de modéliser le système de lentille avec  $N$  plans, la complexité de l’opération de traçage de rayons inverse à plans multiples serait en  $\mathcal{O}(Nn^2)$ . Comparativement, l’algorithme de traçage de rayons inverse à plan unique a une complexité en  $\mathcal{O}(n^2)$ . Pour la modélisation détaillée d’effets de ligne de visée dans des lentilles galaxie-galaxie typiques,  $N$  est généralement autour de 100. Dans des contextes nécessitant un nombre de simulations montant dans les centaines de milliers, voir les millions (par exemple des problèmes d’inférence Bayésienne), la complexité du traçage de rayons inverse à plans multiples devient excessivement prohibitive, malgré le fait que cette modélisation détaillée soit nécessaire aux efforts futurs de l’étude des halos légers de matière sombre.



# Chapitre 4

---

## Apprentissage automatique

The numbers don't decide  
Your system is a lie

---

THOM YORKE

Loin d'être une qualité d'un androïde au coeur de circuits froids qui se voit contempler le but de son existence synthétique, on utilisera ici le terme intelligence artificielle (IA) pour décrire la discipline académique de la théorie et du développement de modèles mathématiques et statistiques pour la résolution de problèmes. Par exemple, un système décisionnel basé sur une séquence de règles précises pour répondre à une certaine question tomberait sous l'égide de l'IA. Ce champ est intimement lié à celui de la science des données, la construction générale de modèles d'IA étant faite sur des machines digitales comme des ordinateurs capable du traitement efficace de données. Le terme apprentissage automatique réfère à une branche de l'IA où l'entraînement de modèles leur permet de développer leurs propres "règles" par l'extraction de motifs et régularités des données, apprenant à généraliser par l'expérience. Ce genre de modèle est construit par une hiérarchie de concepts simples permettant la compréhension et la représentation de concepts complexes. La visualisation de tels modèles comme des graphes de concepts interreliés en plusieurs couches de complexité, ce que l'on appelle réseaux de neurones, décrit l'approche connue comme l'apprentissage profond. Ce chapitre présente en section 4.1 une brève introduction aux modèles linéaires ainsi qu'une description des éléments constitutifs des réseaux de neurones. La section 4.2 décrit les principes derrière l'entraînement de tels réseaux. Enfin, la section 4.3 présente le mécanisme d'attention et les réseaux autoattentifs d'ensembles.

## 4.1. Fondements de l'apprentissage automatique

Il sera d'abord utile d'établir la notion clé d'un espace de données. Supposons qu'on ait un jeu de données  $\mathcal{D} \in \mathbb{R}^d$  dans l'espace des images RGB de  $32 \times 32$  pixels. La résolution en pixels de ces images et leur nombre de canaux de couleur détermine la dimension de cet espace, dans ce cas-ci  $\mathbb{R}^d = \mathbb{R}^{3 \times 32 \times 32} = \mathbb{R}^{3072}$ . Un exemple  $\mathbf{x} \in \mathcal{D}$  tiré de ce jeu de données est un vecteur correspondant à un point en cet espace de très haute dimension. Hors, si  $\mathcal{D}$  rassemble des images "naturelles" d'objets, de scènes ou de sujets vraisemblables, la chance qu'un point tiré au hasard dans  $\mathbb{R}^{3072}$  ressemble en quoi que ce soit à une image qui pourrait appartenir à  $\mathcal{D}$  est infime; il aurait probablement l'allure d'une image sur une télévision à tubes cathodiques aux antennes mal alignées. En fait, l'hypothèse de la variété (*manifold* en anglais) veut que  $\mathcal{D}$  existe sur une variété latente de basse dimension dans l'espace de haute dimension  $\mathbb{R}^d$ . Par exemple, le sous-espace où  $\mathcal{D}$  existe favoriserait sûrement les corrélations spatiales entre les pixels adjacents. La motivation derrière l'apprentissage automatique est de développer des modèles capable d'apprendre à naviguer et organiser les variétés complexes de ces espaces de données afin d'en extraire les informations leur permettant d'accomplir des tâches. J'invite le lecteur à consulter les manuels de Bishop et Nasrabadi (2006) et Goodfellow *et al.* (2016) pour d'excellentes discussions sur ces sujets.

### 4.1.1. Modèles linéaires

Considérons un problème de régression simple, où l'on connaît les valeurs  $y^{(1:N)}$  d'une fonction inconnue  $f(\mathbf{x}) : \mathbb{R}^d \rightarrow \mathbb{R}$  à un échantillon de points  $\mathbf{x}^{(1:N)}$ , avec  $y \in \mathbb{R}$  et  $\mathbf{x} \in \mathbb{R}^d$ . On souhaite modéliser cette fonction avec une famille paramétrique  $g_\theta$  afin d'estimer ses valeurs à des nouveaux points. Une des approches les plus simples pour ce genre de problème serait celle d'un modèle linéaire

$$\begin{aligned}\hat{y}(\mathbf{x}; \mathbf{w}) &= g_{\mathbf{w}}(\mathbf{x}) \\ &= \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x}) \\ &= \mathbf{w}^\top \boldsymbol{\phi}(\mathbf{x})\end{aligned}\tag{4.1.1}$$

où la notation  $\hat{y}$  représente un estimé de la cible  $y$  pour un point  $\mathbf{x}$ . Les poids  $\mathbf{w} = (w_0, \dots, w_{M-1})^\top$  pondèrent différentes fonctions de base  $\boldsymbol{\phi} = (\phi_0, \dots, \phi_{M-1})^\top$ , qui peuvent être des fonctions non-linéaires des entrées  $\mathbf{x}$ . Par exemple, le cas univarié où  $\mathbf{x} = x \in \mathbb{R}$  et  $\phi_j(x) = x^j$  correspond à un modèle de régression polynomiale d'ordre  $M$ . Il est courant de définir  $\phi_0 = 1$  et d'identifier le poids  $w_0$  comme étant un paramètre de biais au modèle. Bien que les fonctions de base  $\phi_j$  soient

non-linéaires en  $\mathbf{x}$ , on réfère tout de même aux fonctions de la forme (4.1.1) comme des modèles linéaires de par leur relation linéaire entre  $\phi$  et  $y$ . Ces modèles peuvent être ajustés à un jeu de données  $\mathbf{x}^{(1:N)}$  en trouvant les poids  $\mathbf{w}$  qui minimisent une certaine fonction d'erreur, ou fonction objectif. Le choix canonique de fonction d'erreur est la moyenne des différences carrées

$$\mathcal{L}(\mathbf{x}, \mathbf{y}; \mathbf{w}) = \frac{1}{2} \frac{1}{N} \sum_{i=1}^N \left\| \hat{\mathbf{y}}^{(i)}(\mathbf{x}^{(i)}; \mathbf{w}) - \mathbf{y}^{(i)} \right\|_2^2, \quad (4.1.2)$$

écrite ci-haut en toute généralité, pour des entrées  $\mathbf{x}$  de dimension arbitraire,  $\mathbf{y} \in \mathbb{R}^M$  et  $\|\cdot\|_2$  la norme  $\ell^2$ . Comme cette fonction d'erreur est quadratique en  $\mathbf{w}$ , sa dérivée par rapport aux paramètres est linéaire en  $\mathbf{w}$ . Cela implique qu'il existe une fonction  $g_{\mathbf{w}^*}$  correspondant à l'unique minimum de  $\mathcal{L}$  qui peut être trouvée analytiquement. Cette propriété est particulièrement utile d'un point de vue d'optimisation, et l'utilisation de fonctions non-linéaires des entrées  $\mathbf{x}$  confère à cette famille de modèles un certain degré d'expressivité, tout dépendant de  $M$ , le nombre de fonctions de base utilisées.

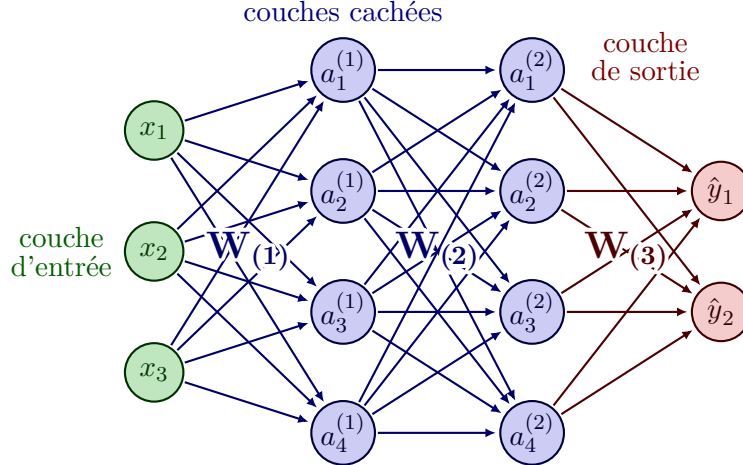
Cependant, la limitation principale des modèles linéaires provient du fait que les fonctions de base  $\phi_j(\mathbf{x})$  sont fixées indépendamment des données  $\mathbf{x}$ . Dans la plupart des applications,  $\mathbf{x}$  est de très haute dimension. La taille de  $M$  doit donc croître exponentiellement pour capturer toutes les interactions entre différents points et généraliser adéquatement. Il s'agit là d'une manifestation du fléau de la dimension, où les points de données deviennent de plus en plus épars et isolés à mesure que leur nombre de caractéristiques — c.-à-d. la dimension de l'espace de données — augmente.

#### 4.1.2. Réseaux de neurones à propagation avant

Héritiers des perceptrons, des modèles mathématiques développés au 20<sup>e</sup> siècle cherchant à représenter le traitement d'information dans les systèmes biologiques (Rosenblatt, 1961), les réseaux de neurones tels qu'on les connaît aujourd'hui sont une classe de familles de fonctions paramétriques qui évitent les lacunes principales des modèles linéaires en apportant une modification importante à la forme (4.1.1): ils admettent des fonctions de base  $\phi$  qui dépendent de paramètres ajustables à un jeu de données  $\mathcal{D} = \{\mathbf{x}^{(1:N)}\}$ .

D'abord, il est utile de considérer la forme mathématique d'un perceptron. Celui-ci agit comme un classificateur binaire, qui a pour forme

$$g_{\mathbf{w}}(\mathbf{x}) = H(\mathbf{w}^\top \mathbf{x} + b) \quad (4.1.3)$$



**Fig. 4.1.** Schéma d'un réseau de neurones à trois couches. Dans ce choix de nomenclature, "trois couches" fait référence aux deux couches cachées avec paramètres  $\mathbf{W}_{(1)}$  et  $\mathbf{W}_{(2)}$  ainsi qu'à la couche de sortie  $\mathbf{W}_{(3)}$ . La couche d'entrée n'est pas comptée dans le nombre de couches.

où  $H(\cdot)$  est la fonction de Heaviside, retournant 0 pour un argument négatif et 1 pour un argument positif. Le perceptron sous sa forme originale ne pouvait que résoudre des ensembles de données  $\mathcal{D}$  étant linéairement séparables. C'est cette importante limitation qui motivera le développement de familles de fonctions plus expressives, comme les réseaux de neurones. La forme archétypique d'un réseau de neurones est une composition de plusieurs fonctions en couches, avec chacune d'entre-elles rappelant celle d'un perceptron. Par exemple, le réseau à 3 couches illustré en figure 4.1 s'écrirait  $g_{\mathbf{w}}(\mathbf{x}) = (g^{(3)} \circ g^{(2)} \circ g^{(1)})(\mathbf{x})$ , où la notation  $g^{(i)}$  identifie la fonction de la  $i$ -ème couche du réseau. Cette forme simple de réseau de neurones prend le nom de **perceptron multicouche**. Ces modèles sont désignés comme étant à **propagation avant**, puisque l'information se propage depuis les entrées  $\mathbf{x}$  par les couches  $g^{(i)}$  jusqu'à la sortie  $\hat{\mathbf{y}}$  sans connection inverse, leur donnant la forme de graphes orientés acycliques. On appelle neurones les unités qui construisent les couches de ces réseaux.

Dans le réseau de la figure 4.1, les neurones de la première couche opèrent d'abord des combinaisons linéaires de toutes les composantes du vecteur d'entrée  $\mathbf{x} \in \mathbb{R}^d$ , produisant les préactivations

$$h_k^{(1)} = \sum_{j=1}^d w_{k,j}^{(1)} x_j + b_k^{(1)} \quad (4.1.4)$$

où  $j = 1, \dots, d$  identifie les composantes de  $\mathbf{x}$  et  $k = 1, \dots, K_1$  identifie les neurones. On reconnaît encore une fois les poids  $w_{k,j}^{(1)}$  et les biais  $b_k^{(1)}$ , qui sont les paramètres de ces neurones. Les activations  $a_k^{(1)}$  des neurones sont ensuite obtenues en passant ces préactivations dans une fonction non-linéaire

différentiable agissant indépendamment sur chaque neurone, appelée fonction d'activation:

$$a_k^{(1)} = \sigma(h_k^{(1)}) \quad (4.1.5)$$

En pratique, une composante  $x_0 = 1$  est ajoutée au vecteur entrant dans la couche, afin de pouvoir regrouper le biais d'un neurone parmi les poids et le faire émerger par le produit  $w_{k,0}^{(1)}x_0 = b_k^{(1)}$ . Les paramètres de la première couche peuvent ainsi être organisés en forme de matrice  $\mathbf{W}_{(1)} \in \mathbb{R}^{(d+1) \times K_1}$ , permettant d'écrire l'opération entière de la première couche comme

$$\mathbf{a}^{(1)} = \sigma(\mathbf{W}_{(1)}^\top \mathbf{x}) \quad (4.1.6)$$

Jusqu'à présent, la forme atteinte en (4.1.6) ressemblant à  $\mathbf{a}^{(1)} = \phi(\mathbf{x})$  n'est pas encore exactement comparable à celle de (4.1.1). C'est lorsqu'on ajoute une seconde couche que l'on remarque leur similitude, et qu'émerge une propriété fondamentale à l'apprentissage automatique, qui sera discutée plus loin. De façon semblable à la première couche, chaque neurone de la seconde couche équivaut à une combinaison linéaire de chaque élément de la couche précédente, suivie d'une activation non-linéaire. Le tout s'écrit

$$\begin{aligned} \mathbf{a}^{(2)} &= \sigma(\mathbf{W}_{(2)}^\top \mathbf{a}^{(1)}) \\ &= \sigma(\mathbf{W}_{(2)}^\top \sigma(\mathbf{W}_{(1)}^\top \mathbf{x})) \end{aligned} \quad (4.1.7)$$

où  $\mathbf{W}_{(2)} \in \mathbb{R}^{(K_1+1) \times K_2}$ , une composante  $a_0^{(1)} = 1$  ayant encore une fois été introduite. La forme entière du réseau de la figure 4.1 serait

$$\begin{aligned} \hat{y}(\mathbf{x}; \mathbf{w}) &= \mathbf{h}^{(3)}(\mathbf{a}^{(2)}(\mathbf{h}^{(2)}(\mathbf{a}^{(1)}(\mathbf{h}^{(1)}(\mathbf{x})))) \\ &= \mathbf{W}_{(3)}^\top \sigma(\mathbf{W}_{(2)}^\top \sigma(\mathbf{W}_{(1)}^\top \mathbf{x})) \end{aligned} \quad (4.1.8)$$

avec  $\mathbf{W}_{(3)} \in \mathbb{R}^{(K_2+1) \times K_3}$ , et  $\mathbf{w}$  l'ensemble de tous les poids du modèle. Il est également possible d'ajouter une fonction d'activation après la couche de sortie, ce qui est commun pour des tâches de classification. Le choix de fonction d'activation interne  $\sigma$ , du nombre de neurones par couches (**largeur**) et du nombre de couches (**profondeur**) sont tous des aspects faisant partie du choix d'**architecture** d'un réseau de neurones. Voici quelques exemples de fonctions d'activation  $\sigma(x)$  typiques, toutes définies avec des dérivées continues:

$$\text{sigmoid}(x) = (1 + e^{-x})^{-1}, \quad \text{ReLU}(x) = \begin{cases} x & \text{si } x > 0 \\ 0 & \text{sinon.} \end{cases}, \quad \text{ELU}(x) = \begin{cases} x & \text{si } x > 0 \\ a(e^x - 1) & \text{sinon.} \end{cases}$$

où  $a$  est un hyper-paramètre à être déterminé pour un problème donné.

Sous la forme (4.1.8), il se voit que les réseaux de neurones sont définis par des fonctions de base pondérées qui dépendent de façon non-linéaire des paramètres et des données, ce qui rend ces fonctions de base adaptives. Cette propriété confère aux réseaux de neurones un important avantage sur les modèles linéaires par rapport au fléau de la dimensionalité. Leurs fonctions de base adaptives peuvent exploiter l’hypothèse de la variété, en adaptant leurs paramètres selon une fonction objectif comme (4.1.2) de sorte à ce que les régions de l’espace de données de haute dimension où varient les fonctions de base correspondent à la variété de cet ensemble de données (Bishop et Nasrabadi, 2006). Cependant, la dépendance non-linéaire de ces modèles en leurs paramètres  $\mathbf{w}$  rend les fonctions objectif non-convexes et sans solution analytique, signifiant que les fonctions optimales trouvées en pratique correspondent généralement à des minima locaux de la fonction objectif.

Ce compromis entre une famille de fonctions extrêmement flexibles et le problème d’optimisation difficile qui leur est associé est souvent justifié par le théorème d’approximation universelle (Cybenko, 1989; Hornik *et al.*, 1989). Ce théorème stipule qu’un réseau de neurones  $g_{\mathbf{w}}$  à une couche cachée<sup>1</sup> et des fonctions d’activation non-linéaires peut approximer une fonction continue  $f$  sur un domaine compact à un degré de précision arbitraire, à condition qu’il soit doté de suffisamment de neurones. L’existence d’une telle fonction optimale  $g_{\mathbf{w}^*}$  est garantie par ce théorème, mais il ne décrit pas comment trouver ses paramètres  $\mathbf{w}^*$ . À cet effet, il est nécessaire d’utiliser des méthodes d’optimisation séquentielles à base de gradients pendant la phase d’**entraînement** d’un modèle.

---

<sup>1</sup>c.-à-d. un réseau à deux couches selon la nomenclature utilisée dans ce texte

## 4.2. Entraînement d'un modèle

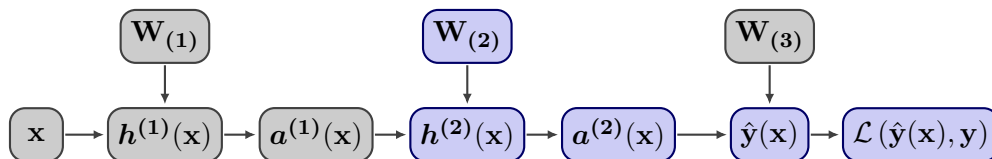
Dans le contexte du développement d'un modèle d'apprentissage automatique  $g_{\mathbf{w}}$  pour le problème de régression présenté en section 4.1.1, on se munit d'un jeu de données d'entraînement consistant d'entrées  $\mathbf{x}^{(1:N)}$  et de cibles associées  $\mathbf{y}^{(1:N)}$ . Comme ce problème comporte des cibles, on dit qu'il s'agit d'un problème **supervisé**. Une métrique de performance du modèle, telle que la moyenne des différences carrées (4.1.2), est employée comme fonction objectif à minimiser dans un processus d'optimisation, que l'on appelle l'entraînement du modèle. Les paramètres  $\mathbf{w}^*$  correspondant à un minimum de cette fonction  $\mathcal{L}$  sont trouvés de façon itérative par un algorithme de descente par gradient, où les poids à une étape  $t - 1$  sont mis à jour en suivant la direction de variation négative de la fonction objectif:

$$\mathbf{w}^{(t)} \leftarrow \mathbf{w}^{(t-1)} - \eta \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{x}, \mathbf{y}; \mathbf{w}^{(t-1)}) \quad (4.2.1)$$

où le paramètre  $\eta > 0$  est le taux d'apprentissage, généralement inférieur à 1. La dépendance de cette méthode sur le gradient de la fonction objectif p.r.a. paramètres du modèle est ce qui motive la construction des réseaux de neurones à partir de fonctions aux dérivées simples, comme des produits matriciels et des fonctions d'activation. Cependant, il est aussi nécessaire d'avoir une méthode efficace de calculer et d'évaluer les gradients d'une fonction objectif p.r.à tous les paramètres d'un modèle.

### 4.2.1. Rétropropagation des gradients

Le passage avant d'un exemple dans la structure d'un réseau de neurones construit un graphe de calcul, identifiant les opérations reliant toutes les composantes du modèle et leurs dépendances fonctionnelles. Par exemple, la figure 4.2 illustre le graphe de calcul du modèle de la figure 4.1. Le gradient de  $\mathcal{L}$  p.r.a. paramètres du réseau peut être facilement obtenu par la règle de dérivation en chaîne en se référant au graphe de calcul, et ses composantes s'expriment sous des formes simples grâce au calcul matriciel. On considère la mise à jour des paramètres  $\mathbf{W}_{(2)}$  de la seconde couche cachée du réseau de la figure 4.1, avec une fonction objectif de la forme (4.1.2) évaluée à un seul point de données  $\mathbf{x}$ . La forme de la dérivée en chaîne s'obtient en passant à reculons au travers de



**Fig. 4.2.** Graphe de calcul associé au réseau de neurone de la figure 4.1. Les vertex impliqués dans le gradient de  $\mathcal{L}$  p.r.à  $\mathbf{W}_{(2)}$  sont identifiés en bleu.

tous les vertex du graphe en aval de celui de la quantité d'intérêt, tel qu'identifiés en figure 4.2. Le gradient de  $\mathcal{L}$  p.r.à ces poids prendrait donc la forme

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}_{(2)}} = \frac{\partial \mathcal{L}}{\partial \hat{\mathbf{y}}} \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{a}^{(2)}} \frac{\partial \mathbf{a}^{(2)}}{\partial \mathbf{h}^{(2)}} \frac{\partial \mathbf{h}^{(2)}}{\partial \mathbf{W}_{(2)}}. \quad (4.2.2)$$

On trouve les expressions des dérivées individuelles:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \hat{\mathbf{y}}} &= \frac{1}{2} \frac{\partial}{\partial \hat{\mathbf{y}}} \|\hat{\mathbf{y}} - \mathbf{y}\|_2^2 \\ &= \hat{\mathbf{y}}(\mathbf{x}) - \mathbf{y} \end{aligned} \quad (4.2.3)$$

$$\begin{aligned} \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{a}^{(2)}} &= \frac{\partial}{\partial \mathbf{a}^{(2)}} \mathbf{W}_{(3)}^\top \mathbf{a}^{(2)} \\ &= \mathbf{W}_{(3)}^\top \end{aligned} \quad (4.2.4)$$

$$\begin{aligned} \frac{\partial \mathbf{a}^{(2)}}{\partial \mathbf{h}^{(2)}} &= \frac{\partial}{\partial \mathbf{h}^{(2)}} \sigma(\mathbf{h}^{(2)}) \\ &= \sigma'(\mathbf{h}^{(2)}(\mathbf{x})) \end{aligned} \quad (4.2.5)$$

La 4<sup>e</sup> dérivée est celle d'un vecteur p.r.à une matrice, nécessitant un traitement plus méticuleux. On procède en notation indicielle, dérivant un élément du vecteur par un élément de la matrice:

$$\begin{aligned} \frac{\partial (\mathbf{h}^{(2)})_i}{\partial (\mathbf{W}_{(2)})_{mn}} &= \frac{\partial}{\partial w_{mn}^{(2)}} \left( \mathbf{W}_{(2)}^\top \mathbf{a}^{(1)} \right)_i \\ &= \frac{\partial}{\partial w_{mn}^{(2)}} \left( w_{ji}^{(2)} a_j^{(1)} \right) \\ &= \delta_{jm} \delta_{in} a_j^{(1)} \\ &= \delta_{in} a_m^{(1)} \\ &= \left( \mathbf{J}^{nm} \mathbf{a}^{(1)} \right)_i \end{aligned} \quad (4.2.6)$$

où la convention de sommation d'Einstein est employée pour les indices répétés et  $\delta$  est le delta de Kronecker. La relation atteinte à la dernière ligne ci-haut (Petersen *et al.*, 2008, éq. 75) introduit la matrice à entrée unique  $\mathbf{J}^{nm} \in \mathbb{R}^{K_2 \times (K_1+1)}$ , avec valeur 1 en  $(n, m)$  et 0 ailleurs. La présence de 3 indices libres dans (4.2.6) indique que le résultat de cette dérivée est un tenseur d'ordre 3:

$$\frac{\partial \mathbf{h}^{(2)}}{\partial \mathbf{W}_{(2)}} = \mathcal{T}^{(2)} \left( \mathbf{a}^{(1)}(\mathbf{x}) \right) \in \mathbb{R}^{K_2 \times (K_1+1) \times K_2} \quad (4.2.7)$$

qui correspond simplement à une réorganisation des composantes de  $\mathbf{a}^{(1)}$  dans un tenseur épars, prenant des valeurs non-nulles seulement où il y a dépendance entre une composante de  $\mathbf{h}^{(2)}$  et un poids de la matrice  $\mathbf{W}_{(2)}$ . Revenant au produit matriciel de dérivées (4.2.2), on y accumule



le gradient en commençant à la fin du graphe de calcul et en remontant le courant, opérant les produits de gauche à droite. En y rassemblant (4.2.3), (4.2.4), (4.2.5) et (4.2.7),

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mathbf{W}_{(2)}} &= (\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{y}) \mathbf{W}_3^\top \odot \sigma'(\mathbf{h}^{(2)}(\mathbf{x})) \mathcal{T}^{(2)}(\mathbf{a}^{(1)}(\mathbf{x})) \\ &= \mathcal{T}^{(2)}(\mathbf{a}^{(1)}(\mathbf{x})) \times_1 \left[ (\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{y}) \mathbf{W}_3^\top \odot \sigma'(\mathbf{h}^{(2)}(\mathbf{x})) \right]\end{aligned}\quad (4.2.8)$$

où  $\odot$  est le produit de Hadamard<sup>2</sup> et  $\times_1$  est le produit mode-1<sup>3</sup> entre le tenseur  $\mathcal{T}^{(2)}$  et le vecteur entre parenthèses carrées, ce qui est l'équivalent du produit matriciel pour des tenseurs selon une dimension spécifiée du tenseur (la première dans ce cas-ci). Le résultat du produit (4.2.8) est une matrice de dimension  $\mathbb{R}^{(K_1+1) \times K_2}$ , c.-à-d. la même dimension que la matrice  $\mathbf{W}_{(2)}$ . Ce gradient peut alors être utilisé dans l'équation (4.2.1) pour mettre à jour les paramètres de la seconde couche du réseau. Les gradients de la fonction objectif par rapport aux paramètres des autres couches du modèle s'obtiennent de la même façon, c.-à-d. en remontant aux vertex associés dans le graphe de calcul 4.2.

L'algorithme de rétropropagation des gradients tire avantage du fait que les gradients de la fonction objectif dépendent principalement des activations et préactivations des neurones du réseau. En conservant en mémoire ces quantités obtenues lors de la propagation avant d'exemples dans le réseau, les gradients p.r.a. paramètres sont facilement obtenus par la propagation arrière des gradients depuis la fonction objectif, selon des opérations dictées par un graphe de calcul. Or, on voit par l'exemple ci-haut qu'il devient rapidement encombrant de trouver manuellement l'expression de ces dérivées en chaînes, surtout pour des réseaux de neurones profonds. C'est en cet aspect que la méthode de **dérivation automatique** implémentée dans des bibliothèques d'apprentissage automatique telles que PyTorch (Paszke *et al.*, 2019) et TensorFlow (Abadi *et al.*, 2015) devient particulièrement utile. Cette approche exploite le fait que les fonctions utilisées dans les réseaux de neurones peuvent être réduites à des compositions d'opérations arithmétiques simples. Par l'application répétée de la règle de dérivation en chaîne, il est ainsi possible d'obtenir la dérivée exacte de fonctions complexes. Ces bibliothèques permettent également la construction automatisée du graphe de calcul, et sont optimisées pour les opérations vectorielles en hautes dimensions.

<sup>2</sup>Le produit de Hadamard, aussi appelé produit élément-par-élément, opère comme suit: pour deux matrices  $A, B \in \mathbb{R}^{m \times n}$ , le produit de Hadamard  $A \odot B \in \mathbb{R}^{m \times n}$  a les composantes  $(A \odot B)_{ij} = (A)_{ij}(B)_{ij}$ . Ce produit apparaît parce que  $\sigma(\cdot)$  est une fonction élément-par-élément.

<sup>3</sup>Le produit mode- $n$  d'un tenseur  $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  avec une matrice  $U \in \mathbb{R}^{J \times I_n}$  s'écrit  $\mathcal{T} \times_n U$  et est de dimensions  $I_1 \times \dots \times I_{n-1} \times J \times I_{n+1} \times \dots \times I_N$ . Ses composantes s'expriment comme

$$(\mathcal{T} \times_n U)_{i_1 \dots i_{n-1} j i_{n+1} \dots i_N} = \sum_{i_n=1}^{I_n} \tau_{i_1 i_2 \dots i_N} u_{j i_n}$$

### 4.2.2. Algorithmes d'optimisation

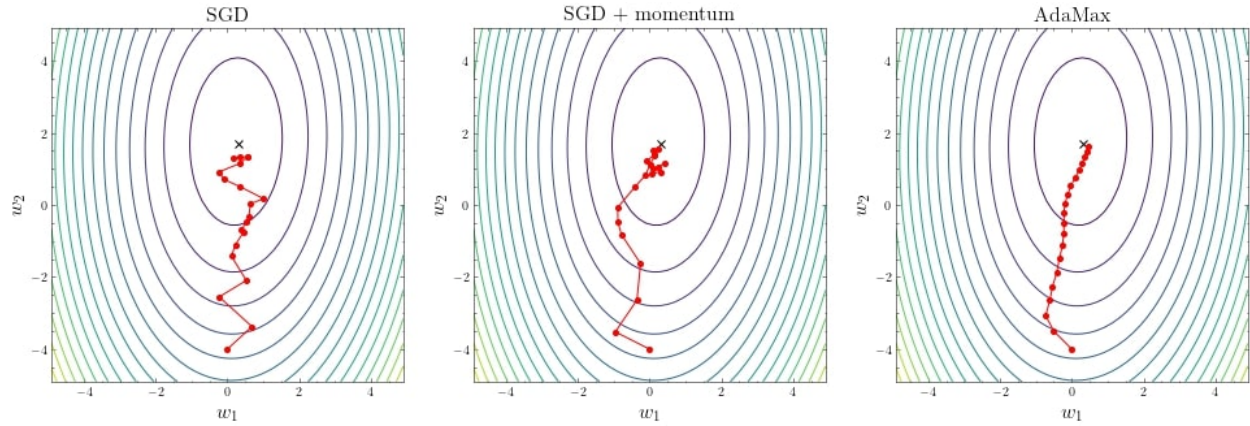
On note que la fonction objectif dans (4.2.1) est évaluée pour le lot du jeu de données. En raison de limitations computationnelles prévenant l'évaluation simultanée de tous les exemples ainsi qu'à des fins de dynamique d'entraînement, il est généralement avantageux d'effectuer des mises à jour des paramètres avec des estimés de la fonction d'erreur du jeu de données entier. En pratique, cette fonction est évaluée pour des sous-ensembles de points de données choisis au hasard, appelés mini-lots (*mini-batch* en anglais). Cela donne lieu à la **descente par gradient stochastique** (SGD), où la mise à jour des poids s'obtient à partir du gradient de la moyenne de la fonction objectif pour un mini-lot  $b = \mathbf{x}^{(1:m)}$  de taille  $m$ :

$$\mathbf{w}^{(t)} \leftarrow \mathbf{w}^{(t-1)} - \eta \nabla_{\mathbf{w}} \mathbb{E}_b \left[ \mathcal{L}(\mathbf{x}, \mathbf{y}; \mathbf{w}^{(t-1)}) \right] \quad (4.2.9)$$

Bien qu'il utilise une approximation du véritable gradient, cet algorithme permet généralement de converger à une petite valeur de la fonction objectif. Cependant, un choix judicieux du taux d'apprentissage  $\eta$  est nécessaire pour atteindre une dynamique de convergence adéquate. Il est commun d'employer un régime de variation de  $\eta$ , débutant généralement à une valeur élevée et diminuant selon un certain profil au cours de l'entraînement. La limite extrême du SGD est celle d'un mini-lot d'un seul exemple, ce qui ralentit généralement la convergence en raison des déplacements erratiques dans l'espace de paramètres que cela entraîne. Un mini-lot comprenant quelques dizaines d'exemples est préférable, puisque le moyennage des gradients mène à des déplacements plus lisses vers un minimum de la fonction objectif.

Populaire pour sa simplicité, l'algorithme SGD est vastement utilisé en apprentissage automatique. Toutefois, il souffre de plusieurs désavantages, notamment son comportement dans des régions irrégulières de l'espace de paramètres. Heureusement, il existe une littérature très riche sur l'optimisation en apprentissage automatique, offrant une panoplie d'algorithmes plus sophistiqués basés sur SGD et visant à améliorer l'optimisation dans les espaces en hautes dimensions. La méthode du **momentum** (Polyak, 1964) s'inspire du concept physique de l'impulsion, en introduisant la quantité de vitesse  $\mathbf{v}$ . Cette dernière représente la direction et vitesse à laquelle une "particule" se déplace à travers l'espace de paramètres. La forme de la règle de mise à jour consiste en une combinaison linéaire des gradients aux pas précédents par décroissance exponentielle:

$$\begin{aligned} \mathbf{v}^{(t)} &\leftarrow \alpha \mathbf{v}^{(t-1)} - \eta \nabla_{\mathbf{w}} \mathbb{E}_b \left[ \mathcal{L}(\mathbf{x}, \mathbf{y}; \mathbf{w}^{(t-1)}) \right] \\ \mathbf{w}^{(t)} &\leftarrow \mathbf{w}^{(t-1)} + \mathbf{v}^{(t)} \end{aligned} \quad (4.2.10)$$



**Fig. 4.3.** Courbes de niveau de la fonction objectif (4.1.2) et trajectoires de différents algorithmes d’optimisation dans l’espace des paramètres à deux dimensions d’un problème de régression linéaire. Chacun des panneaux illustre 20 pas d’entraînement à partir des mêmes paramètres initiaux, pour des mini-lots de 2 exemples. Les courbes de niveaux de la fonction objectif  $\mathcal{L}$  correspondent aux valeurs de  $\mathcal{L}$  évaluée pour l’ensemble de données entier de 100 exemples.

où  $\alpha \in [0,1)$  est un hyper-paramètre déterminant l’importance que prennent les gradients précédents dans la direction de la mise à jour. Cette inclusion de l’historique des gradients permet un partage de l’information entre les mini-lots pour une mise à jour des poids donnée, accélérant la convergence vers un minimum de la fonction objectif par rapport à SGD.

La performance d’un réseau de neurones est très sensible au taux d’apprentissage  $\eta$ , ce qui en fait l’un des hyper-paramètres les plus difficiles à déterminer. Le régime de réduction graduelle de  $\eta$  mentionné plus tôt présente déjà un net avantage sur un  $\eta$  fixe quant à la convergence d’un modèle vers une bonne solution, malgré que le choix de ce régime représente un hyper-paramètre en soi. C’est pourquoi les algorithmes d’optimisation modernes les plus répandus reposent sur une approche où des taux d’apprentissage peuvent être adaptés automatiquement selon la sensibilité de la fonction objectif aux différents paramètres. L’un des plus utilisés est celui d’**AdaMax** (Kingma et Ba, 2014), qui pondère la mise à jour de paramètres individuels de façon inversement proportionnelle à la norme infinie de leur gradients présents et passés. Cet algorithme incorpore des aspects de SGD et du momentum, par son estimation du moment de premier ordre du gradient. Cette quantité est initialisée à  $\mathbf{m}^{(0)} = 0$ , tout comme la norme infinie  $\mathbf{u}^{(0)} = 0$ . La règle de mise à jour prend la forme

$$\begin{aligned}
\mathbf{g}^{(t)} &\leftarrow \nabla_{\mathbf{w}} \mathbb{E}_b [\mathcal{L}(\mathbf{x}, \mathbf{y}; \mathbf{w}^{(t-1)})] && \text{(gradient de } \mathcal{L} \text{ pour un mini-lot)} \\
\mathbf{m}^{(t)} &\leftarrow \beta_1 \mathbf{m}^{(t-1)} + (1 - \beta_1) \mathbf{g}^{(t)} && \text{(estimé du moment de premier ordre)} \\
\mathbf{u}^{(t)} &\leftarrow \max(\beta_2 \mathbf{u}^{(t-1)}, |\mathbf{g}^{(t)}|) && \text{(mise à jour de la norme infinie)} \\
\mathbf{w}^{(t)} &\leftarrow \mathbf{w}^{(t-1)} - \frac{\eta}{1 - \beta_1^t} \cdot \frac{\mathbf{m}^{(t)}}{\mathbf{u}^{(t)}} && \text{(mise à jour des paramètres)}
\end{aligned} \tag{4.2.11}$$

où les hyper-paramètres  $\beta_1, \beta_2 \in [0, 1)$  sont les taux de décroissance exponentielle du premier moment et de la norme infinie, respectivement. Il s’observe en figure 4.3 que la trajectoire d’optimisation avec AdaMax est bien plus lisse que celles avec SGD ou momentum. Effectivement, les estimés du véritable gradient par ces deux autres algorithmes sont visiblement biaisés par la petite taille des mini-lots, alors qu’Adamax parvient à bien estimer la direction du minimum global de la fonction objectif convexe.

### 4.3. Réseaux autoattentifs

Aussi connus sous le nom de Transformeurs, les réseaux de neurones autoattentifs (Vaswani *et al.*, 2017) révolutionnent depuis quelques années le champ du traitement automatique de la langue naturelle, détrônant les réseaux de neurones récurrents à des tâches comme la traduction et la compréhension du langage. L’exemple récent le mieux connu d’un grand modèle de langage à base de Transformeurs est ChatGPT (OpenAI, 2023). Plus généralement, ce type de modèle excelle à la modélisation de données séquentielles, ayant la capacité de capturer les interactions entre tous les éléments indépendamment de leur séparation à l’intérieur d’une séquence. Cela s’avère à être particulièrement utile à l’extraction de représentations cachées significatives, un aspect crucial pour les capacités d’abstraction et de synthèse des modèles d’apprentissage automatique.

#### 4.3.1. Mécanisme d’attention

Le fonctionnement des réseaux autoattentifs repose sur le mécanisme d’attention, une fonction mettant en correspondance des vecteurs requête (*query*) et des paires de vecteurs clé-valeur (*key-value*). En général, les requêtes sont extraites d’une séquence d’entrée de longueur  $s$ , et les paires clé-valeur d’une autre séquence de longueur  $n_v$ , construisant les matrices  $Q \in \mathbb{R}^{n \times d_q}$ ,  $K \in \mathbb{R}^{n_v \times d_q}$  et  $V \in \mathbb{R}^{n_v \times d_v}$ . La fonction d’attention s’exprime

$$\text{Attention}(Q, K, V; \omega) = \omega \left( QK^\top \right) V \tag{4.3.1}$$

où  $\omega$  est une fonction d'activation, prenant généralement la forme d'un softmax normalisé:  $\omega(\cdot) = \text{softmax}(\cdot / \sqrt{d_q})$ , avec  $\text{softmax}(z) = \frac{e^{z_i}}{\sum_i e^{z_i}}$ . Le résultat de (4.3.1) est une somme pondérée des valeurs  $V$ , avec des poids provenant d'une fonction de compatibilité entre les requêtes  $Q$  et les clés  $K$  associée aux valeurs. Une valeur prendra plus d'importance s'il y a une forte correspondance entre sa clé et la requête dans le produit scalaire  $QK^\top$ . On appelle **autoattention** un scénario où les vecteurs  $Q, K, V$  proviennent tous de la même séquence. Une couche d'autoattention a la capacité de capturer les interactions entre chaque paire d'éléments d'une séquence d'entrée, ce qui représente un atout crucial en modélisation du langage où le sens d'une phrase peut changer si une certaine paire de mots y apparaît.

Vaswani *et al.* (2017) ont introduit l'attention à têtes multiples (*multi-head attention*), une extension marquante du mécanisme d'attention où l'on dénombre  $h$  "têtes" d'attention plutôt qu'une seule. D'abord, les quantités  $Q, K, V$  sont projetées  $h$  fois par des paramètres appris sur des vecteurs aux dimensions respectives  $d_q^M, d_k^M, d_v^M$ . Chacune des  $h$  têtes applique une fonction d'attention indépendante à l'une de ces  $h$  projections, ce qui est finalement concaténé et subit une transformation linéaire:

$$\text{Multihead}(Q, K, V; \theta, \omega) = \text{concat}(O_1, \dots, O_h)W^O, \quad (4.3.2)$$

$$\text{où } O_j = \text{Attention}(QW_j^Q, KW_j^K, VW_j^V; \omega_j) \quad (4.3.3)$$

où  $\theta = \{W_j^Q, W_j^K, W_j^V\}_{j=1}^h$  sont les paramètres appris servant de matrices de projection, avec  $W_j^Q, W_j^K \in \mathbb{R}^{d_q \times d_q^M}$ ,  $W_j^V \in \mathbb{R}^{d_v \times d_v^M}$  et  $W^O \in \mathbb{R}^{hd_v^M \times d}$ . Les choix typiques pour les hyperparamètres dimensionnels sont de  $d_q = d_v = d$ ,  $d_q^M = d_v^M = d/h$ , avec  $d$  la dimension d'un des  $n$  éléments de la séquence. La méthode d'attention à têtes multiples est particulièrement avantageuse puisque les  $h$  têtes peuvent opérer de façon parfaitement parallèle, et permettent d'extraire une variété de différentes représentations d'une même séquence.

### 4.3.2. Réseaux autoattentionnels d'ensembles

Plusieurs applications de l'apprentissage des représentations concernent le traitement de données sous forme d'**ensemble**, un regroupement d'un nombre arbitraire de vecteurs de dimension communes mais où l'ordre n'a pas d'importance. Un modèle capable d'extraire des représentations de dimensions fixes à partir d'entrées structurées en ensembles doit satisfaire deux conditions liées à la définition même d'un ensemble. En premier lieu, ce modèle doit être invariant sous permutations, c.-à-d. que son résultat ne doit pas dépendre de l'ordre des éléments dans l'ensemble d'entrée. D'autre part, il doit avoir la capacité de traiter des ensembles de taille arbitraire.

Lee *et al.* (2019) ont introduit les réseaux autoattentionnels d'ensembles (RAE), un type de modèle à base d'attention satisfaisant les conditions ci-haut. L'architecture d'un RAE suit la forme d'un encodeur-décodeur, où l'encodeur produit un ensemble de représentations cachées à partir de l'ensemble d'entrée de taille  $n$ , et le décodeur agrège cet ensemble de caractéristiques pour produire la sortie désirée. L'avantage d'un RAE par rapport à d'autres modèles d'ensembles est par son architecture d'encodeur en plusieurs couches d'autoattention consécutives et son mécanisme de sous-échantillonnage par autoattention dans le décodeur. Cela lui permet de capturer des interactions d'ordre supérieur entre les éléments d'un ensemble afin d'en produire une meilleure représentation.

Les auteurs définissent certaines couches neuronales adaptées pour la construction d'un RAE. Ayant deux matrices  $X, Y \in \mathbb{R}^{n \times d}$  représentant deux ensembles de vecteurs  $d$ -dimensionnels, le bloc d'attention à têtes multiples (MAB) aux paramètres  $\theta$  est défini comme

$$\text{MAB}(X, Y) = \text{LayerNorm}(H + \text{rFF}(H)), \quad (4.3.4)$$

$$\text{où } H = \text{LayerNorm}(X + \text{Multihead}(X, Y, Y; \theta)), \quad (4.3.5)$$

rFF est un réseau à propagation avant quelconque opérant par rangées, traitant chaque instance de façon identique et indépendante, et LayerNorm est la normalisation de couche (Ba *et al.*, 2016). Cette normalisation opère sur chaque élément d'un ensemble pour standardiser ses caractéristiques, et est couramment utilisée dans les modèles d'autoattention. Le bloc (4.3.4) sert à définir le bloc d'autoattention d'ensembles (SAB)

$$\text{SAB}(X) = \text{MAB}(X, X). \quad (4.3.6)$$

Ce bloc opère l'autoattention entre les éléments d'un ensemble d'entrée, produisant un ensemble de taille égale. C'est par l'enchaînement de plusieurs couches de SAB qu'un RAE extrait d'un ensemble ses interactions d'ordre supérieur.

Les auteurs remarquent que la complexité temporelle en  $\mathcal{O}(n^2)$  du SAB peut devenir prohibitive dans le cas d'ensembles de grande taille  $n$ . Pour y remédier, ils introduisent le bloc d'autoattention d'ensemble induite (ISAB). Cette modification au SAB nécessite un ensemble de taille  $m < n$  de vecteurs  $d$ -dimensionnels  $I \in \mathbb{R}^{m \times d}$ , appelés points d'induction. Contrairement à un ensemble d'entrée  $X$ , ces points font partie du ISAB même et sont des paramètres à entraîner. Un bloc ISAB

avec  $m$  points d'induction est défini comme

$$\text{ISAB}_m(X) = \text{MAB}(X, H) \in \mathbb{R}^{n \times d}, \quad (4.3.7)$$

$$\text{où } H = \text{MAB}(I, X) \in \mathbb{R}^{m \times d}. \quad (4.3.8)$$

Plutôt que d'opérer une véritable autoattention entre  $X$  et lui-même, ce bloc applique d'abord l'attention à têtes multiples entre  $X$  et le plus petit ensemble de points d'induction  $I$  pour produire  $H$ . Cet ensemble de caractéristiques encode de l'information au sujet de l'entrée  $X$ , et est utilisé dans une seconde application de l'attention avec  $X$  pour produire un ensemble de représentations de taille  $n$ . Il est raisonnable de croire que les points d'induction  $I$  servent à extraire de l'information globale de  $X$  pour en produire la quantité de dimension réduite  $H$ , dont la seconde attention avec  $X$  permet de produire des représentations utiles. Dans le cas du SAB où la complexité  $\mathcal{O}(n^2)$  était quadratique, celle du ISAB est de  $\mathcal{O}(nm)$ . La valeur de  $m$  est un hyper-paramètre à choisir de sorte à ce que l'efficacité du modèle soit améliorée sans sacrifier sa performance.

À travers les couches de l'encodeur d'un RAE composées de SAB ou ISAB consécutifs, l'ensemble de représentations internes demeure de taille  $n$ . C'est à l'arrivée au décodeur que cet ensemble doit subir un certain mécanisme d'agrégation afin de produire une représentation de dimension fixe de l'ensemble d'entrée  $X$ . Les exemples les plus simples de tels mécanismes sont la moyenne ou le maximum par dimension des caractéristiques, des choix couramment employés par d'autres type de modèles d'ensembles. Pour sa part, le mécanisme d'agrégation introduit avec les RAE repose sur une attention à têtes multiples entre l'ensemble de représentations cachées  $Z \in \mathbb{R}^{n \times d}$  issu d'un décodeur et un ensemble appris de  $k$  vecteurs germes  $S \in \mathbb{R}^{k \times d}$ . L'opération d'agrégation par attention à têtes multiples (PMA) avec  $k$  vecteurs germes s'exprime

$$\text{PMA}_k(Z) = \text{MAB}(S, \text{rFF}(Z)). \quad (4.3.9)$$

Le résultat de cette opération est lui-même un ensemble de  $k$  éléments, mais dans la plupart des cas une valeur  $k = 1$  est utilisée pour qu'un RAE produise une seule représentation d'un ensemble  $X$ . L'agrégation de caractéristiques par attention est avantageuse car il est fort probable que l'influence de différents membres d'un ensemble  $Z$  sur l'objectif d'un problème d'apprentissage automatique ne soit pas égale. Le mécanisme PMA peut ainsi être plus attentif à certains éléments qui affectent davantage la cible.

Une architecture typique de RAE pour le traitement d'ensembles  $X$  de grande taille suivant la forme encodeur-décodeur peut être exprimée par une composition des blocs présentés ci-haut:

$$\text{Encodeur}(X) = \text{ISAB}_m(\text{ISAB}_m(X)) \in \mathbb{R}^{n \times d} \quad (4.3.10)$$

$$\text{Décodeur}(Z) = \text{rFF}(\text{SAB}(\text{PMA}_k(Z))) \in \mathbb{R}^{k \times d} \quad (4.3.11)$$

où  $Z$  est la sortie de l'encodeur. Cette architecture représente une structure générale, et fait donc abstraction d'un choix de fonction objectif, qui serait déterminée en fonction du problème d'apprentissage. Un exemple de fonction objectif pour une tâche de régression serait une moyenne des différences carrées, telle (4.1.2), comparant la sortie  $Z$  d'un RAE ayant pris  $X$  en entrée à l'étiquette associée  $y$ .



## Chapitre 5

---

# AlphaGen: a Neural Network for Fast Multi-Plane Lensing

Charles Wilson,<sup>1,2</sup> Yashar Hezaveh,<sup>1,2,3,4</sup> Adam Coogan<sup>1,2,3</sup>

<sup>1</sup>*Département de Physique, Université de Montréal, 1375 Avenue Thérèse-Lavoie-Roux,  
Montréal, QC H2V 0B3, Canada*

<sup>2</sup>*CIELA - Montreal Institute for Astrophysics and Machine Learning, Montréal, Canada*

<sup>3</sup>*Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, 10010, New York,  
NY, USA*

<sup>4</sup>*Mila - Quebec AI Institute, 6666 St-Urbain, #200, Montréal, QC H2S 3H1, Canada*

## Résumé

La modélisation des effets de structure en ligne de visée dans les lentilles gravitationnelles fortes a été historiquement limitée par son importante complexité de calcul. Or, l’analyse de l’immense volume de nouvelles lentilles fortes projetées à être découvertes par l’Observatoire Vera Rubin et le télescope spatial Euclid exige des simulateurs d’effets de ligne de visée plus efficaces que le traçage de rayons à plans multiples, et plus exacts que les approximations à plan unique. Dans cet ouvrage, nous développons un formalisme de modélisation à base de réseaux de neurones servant à la génération d’angles de déflexion multiplans effectifs, conditionnés sur une population de halos en ligne de visée ainsi qu’un profil de lentille principale déterminé par l’usager. Notre pipeline permet une accélération importante de la production de déflexions multiplans, tout en modélisant les effets de milliers de halos en ligne de visée avec exactitude. Les réseaux de neurones qui composent notre pipeline sont compatible avec des entrées sous la forme d’ensembles (des listes de halos de champ de longueurs variables), permettant l’inférence Bayésienne hiérarchique de paramètres de population. Lorsque comparé au formalisme traditionnel de lentillage à plans multiples, notre pipeline accéléré par GPU est plus rapide par presque deux ordres de grandeur à une résolution semblable à celle d’HST, et peut reproduire des lentilles gravitationnelles à plans multiples avec exactitude, ce que l’on démontre par le lentillage d’images de galaxies véritables provenant du recensement COSMOS.

**Mots-clés:** Lentilles gravitationnelles — Matière sombre — Simulations astrophysiques — Réseaux de neurones

## Abstract

Modeling the effects of line-of-sight structure in strong gravitational lenses has historically been limited by its high computational complexity. Yet, the analysis of the immense volume of strong lensing data expected from the Vera Rubin Observatory and the Euclid space telescope calls for simulators of line-of-sight effects that are more efficient than multi-plane ray-tracing, and more precise than single-plane approximations. In this work, we develop a neural-network-based modeling pipeline for the generation of effective multi-plane deflection angles, conditioned on a population of line-of-sight dark matter halos and a user-defined main lens profile. Our pipeline allows to significantly accelerate the production of multi-plane deflections, while accurately capturing the effects of thousands of line-of-sight halos. The pipeline’s neural network components are compatible with set-structured inputs (variable-length lists of interloping halos), allowing for hierarchical Bayesian inference of population-level parameters. When compared to the traditional multi-plane lensing

framework, our GPU-accelerated pipeline is nearly two orders of magnitude faster at an HST-like resolution, and can accurately reproduce multi-plane gravitational lenses, which we demonstrate by lensing real galaxy images from the COSMOS survey.

**Keywords:** Strong gravitational lensing(1643) — Dark matter(353) — Astronomical simulations(1857) — Neural networks(1933)

## 5.1. Introduction

The widely successful  $\Lambda$ CDM paradigm of cosmology is supported by a wealth of cosmological measurements at the largest scales in the universe, be it of the clustering of galaxies, baryonic acoustic oscillations or the cosmic microwave background. However, a number of issues at small scales persist with this model, first among them being the missing satellites problem. This refers to the observed paucity of detected dark matter halos at sub-galactic scales ( $M < 10^{10} M_{\odot}$ ) relative to their predicted abundance from  $\Lambda$ CDM.

This dearth could signify that these halos do not exist, implying a cut-off scale in the halo mass function characteristic of alternative dark matter models such as WDM or SIDM. Conversely, it could just as well be indicative that they have as of yet escaped detection. A number of baryonic processes have been suggested for suppressing star formation efficiency in low-mass halos, indicating that they would harbor too little luminous matter that could serve as a tracer for their detection through observations of satellite galaxies. Constraining dark matter at small scales is difficult, but it is in this low-mass regime that observables can discriminate between different theories and could upheave the standard model of cosmology.

Strong gravitational lensing represents the best probe for the detection of dark matter halos, making it an invaluable tool for discovering the particle properties of dark matter. As lensing only couples to the force of gravity, it is sensitive to distortions in extended arcs introduced by dark substructure within the haloes of massive lens galaxies, also known as subhalos. Galaxy-scale strong lensing has successfully been used in this way to detect individual subhalos in lens galaxies (Vegetti *et al.*, 2010, 2012; Hezaveh *et al.*, 2016) and constrain their position and mass, a method known as gravitational imaging (Koopmans, 2005). The statistical analysis of a sample of lens galaxies by Vegetti *et al.* (2014) has also allowed to place constraints on the subhalo mass function. However, only substructure down to mass scales of  $10^8 - 10^{10} M_{\odot}$  can be directly detected by this approach, with the effect of lighter halos being too faint to be individually resolved.

Another approach relies on quadruply-lensed quasars, a different type of lensing system where the observables are image positions and flux ratios rather than extended arcs. Anomalies between these flux ratios can be associated to the presence of individual subhalos or the combined lensing effect of substructure populations, probing subhalo masses down to  $10^7 M_\odot$ . Past analyses of simulations of such systems have shown them to be promising for constraining the abundance of substructure (Metcalf et Zhao, 2002; Chiba, 2002; Xu *et al.*, 2009; Gilman *et al.*, 2018). Studies of observed lensed quasars have been favorable to models including a single subhalo (Fadely et Keeton, 2012; MacLeod *et al.*, 2013; Nierenberg *et al.*, 2014) and populations of subhalos (Birrer *et al.*, 2017a).

While the works mentioned above have focused on the study of substructure within the main lens, a number of authors have highlighted that intergalactic dark matter halos, i.e. isolated small-scale structure distributed in the line-of-sight (LOS) of a strong lensing system, should have a significant effect on lensing observables and even dominate the substructure signal (Chen *et al.*, 2003; Metcalf, 2005; Amara *et al.*, 2006; Xu *et al.*, 2012; Li *et al.*, 2017; Despali *et al.*, 2018). Neglecting to model the contribution of LOS structure in strong lensing leads to biased inference in areas such as mass modeling, mass function constraints and time-delay cosmography. Intergalactic halos have historically been overlooked in the lensing literature save for recently, e.g. in lensed quasar flux ratio anomalies (Inoue *et al.*, 2015; Hsueh *et al.*, 2020; Gilman *et al.*, 2019, 2020a,b; Gilman *et al.*, 2021; Gilman *et al.*, 2022; Laroche *et al.*, 2022), low-mass halo detection in samples of strong lenses (Ritondale *et al.*, 2019), contribution of LOS halos to the surface mass density power spectrum in strong lenses (Şengül *et al.*, 2020), constraints on dark matter through joint analysis of strong lensing observables and the Milky Way satellite population Nadler *et al.* (2021), distinguishing between LOS halo and substructure signal in the two-point function of the effective lensing deflection field (Dhanasingham *et al.*, 2022), or reanalysis of the subhalo detection of Vegetti *et al.* (2012) to identify the perturber as an interloping extragalactic halo (Şengül *et al.*, 2022).

Bayesian inference methods for data analysis require data modeling with accurate simulators. Generally, the modeling of gravitational lensing by structure in the LOS of a strong lensing system calls for multi-plane lensing (Blandford et Narayan, 1986), an extension of the single-plane lensing framework used for modeling subhalo lensing. However, the multi-plane formalism is much more difficult to treat, as the lens equation describing the deflection of light becomes recursive. This makes the complete modeling of lensing by LOS structure quite computationally expensive, a significant hindrance for inference problems requiring thousands upon thousands of simulations. Motivated in part by this difficulty, other methods have been developed to model the LOS lensing

contribution. Some approximate the matter inhomogeneities in the LOS as tidal perturbations (Kovner, 1987; Bar-Kana, 1996; Schneider, 1997; Birrer *et al.*, 2017a) in the form of added external convergence and shear to the main lens, or construct a scheme for projecting LOS halos to the main lens with rescaled parameters (Şengül *et al.*, 2020). Other works have introduced hybrid frameworks, combining aspects of multi-plane lensing and tidal perturbations (McCully *et al.*, 2014; Schneider, 2014; Fleury *et al.*, 2021a,b). Moreover, Gilman *et al.* (2019) have devised a perturbative algorithm for multi-plane lensing in the background LOS of a main lens galaxy, specifically tailored for the inference of dark matter model parameters with flux ratio statistics. Even so, what these approximate frameworks grant in efficiency, they lack for in accuracy. As noted by McCully *et al.* (2017), multi-plane lensing introduces non-linear effects that cannot be reproduced by single-plane approximations. With regards to hybrid approaches, they remain less accurate than full multi-plane lensing while still introducing some form of slow recursion.

New and upcoming telescopes and surveys such as JWST (Gardner *et al.*, 2006), Euclid (Refregier *et al.*, 2010; Laureijs *et al.*, 2011), ELT (Simon *et al.*, 2019), LSST (LSST Dark Energy Science Collaboration (LSST DESC) *et al.*, 2021) and SKA (Koopmans *et al.*, 2004) are poised to discover hundreds of thousands of new galaxy-galaxy strong lenses (Collett, 2015) in the coming decade and deliver very high-precision imaging in follow-ups, allowing to resolve the impact of matter inhomogeneities on lensing observables. The analysis of this wealth of data will require accurate lensing simulators. However, the immense volume of incoming data will render the analysis of these systems through traditional methods unfeasible on human timescales.

Major advances in machine learning in recent years along with the pressing need for efficient models and inference pipelines have ushered in a new era for the field of strong gravitational lensing. A multitude of novel deep learning methods are proving to be extremely successful at accurate lens modeling (Hezaveh *et al.*, 2017; Perreault Levasseur *et al.*, 2017; Morningstar *et al.*, 2018; Wagner-Carena *et al.*, 2021; Legin *et al.*, 2021; Adam *et al.*, 2023), source reconstruction (Morningstar *et al.*, 2019; Mishra-Sharma et Yang, 2022; Karchev *et al.*, 2022; Adam *et al.*, 2022, 2023), low-mass halo inference (Brehmer *et al.*, 2019; Coogan *et al.*, 2020; Wagner-Carena *et al.*, 2023), time delay cosmography (Park *et al.*, 2021) and population-level inference (Legin *et al.*, 2022; Anau Montel *et al.*, 2023). These developments are laying the groundwork for tackling the analysis of the overwhelming amount of incoming astronomical data, although only a few fully model the LOS lensing contribution. Furthermore, the often intractable likelihoods associated to strong lensing

inference problems has led to a recent surge of interest in the development and use of simulation-based inference methods (Cranmer *et al.*, 2020) in strong lensing. However, the viability of such approaches is critically dependent on fast and accurate simulators.

In this work, we present AlphaGen: a neural network-based pipeline for the acceleration of multi-plane lensing simulations in the context of galaxy-galaxy strong lenses at a resolution comparable to HST. The original network architecture we propose in this work leverages the complete information content of a halo population in the LOS of a galaxy-galaxy strong lensing system to generate the associated deflection angles, all while leaving the halo parameters and main deflector model completely up to the user. Our model then acts as a function producing the equivalent of multi-plane gravitational lensing with two evaluations of feed-forward neural networks, bypassing the need for expensive recursive computations. The implementation of this framework is set to be released as a module in the upcoming strong gravitational lensing software `caustics`.

This paper is organized as follows. Section 5.2 gives an introduction to the mathematical framework of single-plane and multi-plane gravitational lensing. In Section 5.3, we present the architecture of the neural network components of the pipeline, as well as the pipeline’s workflow. In Section 5.4, we detail the procedure for the production of training data. In Section 5.5, we briefly discuss the training strategies for our models. In Section 5.6, we report our results on test halo populations, and present speed benchmarks for our pipeline. We conclude in Section 5.7 and discuss future avenues of development for our pipeline.

## 5.2. Lensing formalism

### 5.2.1. Lens equation and lensing quantities

In the archetypal strong lensing system involving a bright background source at redshift  $z_s$  and a single massive foreground galaxy at  $z_\ell$ , an observed angular position  $\boldsymbol{\theta}$  in the image plane measured relative to the optical axis is related to an angular position  $\boldsymbol{\beta}$  in the source plane by the lens equation

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}), \tag{5.2.1}$$

where  $\boldsymbol{\alpha}$  is the deflection angle. This deflection occurs in a very short span relative to the complete path of a light ray, while it is passing in the vicinity of the mass distribution of the lens galaxy. This justifies the thin lens approximation, where the distribution of matter in the lens is approximated to be projected to a plane. The deflection of light then occurs instantaneously at this *lens plane*,

and only depends on the projected mass density through the integral

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{1}{\pi} \int_{\mathbb{R}^2} \kappa(\boldsymbol{\theta}') \frac{\boldsymbol{\theta} - \boldsymbol{\theta}'}{\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|^2} d^2\boldsymbol{\theta}' \quad (5.2.2)$$

where  $\kappa(\boldsymbol{\theta}) = \Sigma(\boldsymbol{\theta})/\Sigma_c$  is the dimensionless projected surface density, also known as convergence.

It is normalized by a critical surface density

$$\Sigma_c = \frac{c^2}{4\pi G} \frac{D_s}{D_\ell D_{\ell s}}, \quad (5.2.3)$$

a function of angular diameter distances between different segments of the observer-lens( $\ell$ )-source( $s$ ) system. The notation  $D_{ij}$  represents an angular diameter distance to a plane at  $z_j$  seen from  $z_i$ , and where a single subscript (e.g.  $D_s$ ) implies an angular diameter distance seen from the observer plane. In standard strong gravitational lensing frameworks, such cosmological distances are usually defined with respect to a reference Friedmann-Lemaître-Robertson-Walker (FLRW) spacetime, with lenses being perturbers in an otherwise homogeneous Universe.

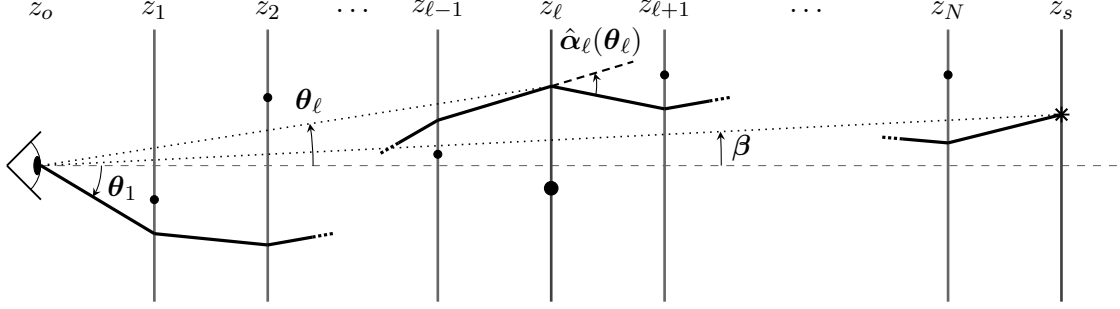
When modeling the effect of strong lenses, the deflected coordinates  $\boldsymbol{\beta}$  are generally obtained through the *backward ray-tracing* algorithm. This procedure involves "shooting" rays from zero-deflection initial conditions at the observer plane, going out at angles  $\boldsymbol{\theta}$  towards a lens plane. These rays are then deflected by the lens at  $z_\ell$  to land on the deflected coordinates  $\boldsymbol{\beta}$  at the source plane, as related through the lens equation (5.2.1).

### 5.2.2. Multiplane lensing

A more complete description of strong gravitational lensing should include the effect of multiple deflections along the line-of-sight, in addition to that of a main lens galaxy. We then model a galaxy-galaxy strong lensing system with interlopers as  $N + 1$  consecutive lens planes orthogonal to the optical axis at redshifts  $z_i$ , where  $i = 1, 2, \dots, N + 1$  such that  $z_j < z_k$  for  $j < k$ . The redshift  $z_o = 0$  denotes the observer plane,  $z_\ell$  the main deflector plane and  $z_s$  the source plane, with  $s \equiv N + 1$ . Such lens planes represent redshift bins of widths defined by the redshift interval between adjacent planes. A given plane can thus hold an arbitrary number of deflectors falling within its redshift bin.

The single-plane lens equation (5.2.1) can be extended to  $N$  planes of deflections to assume the form

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}_{\text{eff}}, \quad \boldsymbol{\alpha}_{\text{eff}} = \sum_{i=1}^N \frac{D_{is}}{D_s} \hat{\boldsymbol{\alpha}}_i(\boldsymbol{\theta}_i) \quad (5.2.4)$$



**Fig. 5.1.** Multi-plane lensing, shown here with a single deflector per lens plane. The dotted horizontal line is the optical axis, and its intersection with each of the  $N + 1$  lens planes is the origin for their angular coordinates  $\theta_i$ . The thick solid line represents the trajectory of a light ray. In backward ray-tracing, the ray comes in through the lens plane at  $z_\ell$  at the angular position  $\theta_\ell$ , and is deflected by an angle  $\hat{\alpha}_\ell$  relative to the plane. This deflection determines the angular position  $\theta_{\ell+1}$  at the next lens plane through eq. (5.2.6).

where  $\alpha_{\text{eff}}$  is the effective deflection angle, compounding the deflections of every lens plane between observer and source. We will refer to the first lens plane ( $i = 1$ ) as the image plane. The notation  $\hat{\alpha}_i$  identifies a "physical" angle, relative to the reference frame of a lens plane  $i$ . The "reduced" angle, relative to the observer plane, is defined as

$$\alpha_i = \frac{D_{is}}{D_s} \hat{\alpha}_i \quad (5.2.5)$$

The deflection field of multiple halos on a single plane is simply the linear addition of deflections from individual halos. However, the same does not hold for the case of multiple lens planes, where deflections at different redshifts become coupled through a recursive equation for the lens plane coordinates:

$$\theta_i = \theta - \sum_{m=1}^{i-1} \frac{D_{mi}}{D_i} \hat{\alpha}_m(\theta_m) \quad (5.2.6)$$

It is by this very recursion that  $\alpha_{\text{eff}}$  depends not only on the image plane coordinates  $\theta$  as is the case in the single-plane lens equation (5.2.1), but on the coordinates at every lens plane. Therein is the foremost efficiency bottleneck of full multi-plane ray-tracing (MPRT) computations, motivating the development of an accelerated framework for MPRT that can bypass this recursion.

### 5.3. The AlphaGen Pipeline

Designing a machine learning model for the simulation of MPRT can be thought of as tackling two separate problems. First, we require a model capable of mapping from a grid of observed angular coordinates to deflection angles, conditioned on a population of dark matter halos in the line-of-sight. Second, the information of such a population should be encoded as a compressed



statistic, summarizing the compounded effect of a multitude of halos on effective deflection angles. Thus, different halo populations would correspond to different encodings, to be used for mapping to their corresponding deflection angles.

In this section, we present the two main components of our model, as well as the end-to-end pipeline for effective deflection angle generation. The Set Model, providing an lower-dimensional encoding of a set of halo parameters, is described in section 5.3.1. The Map Model, whose architecture is presented in section 5.3.2, maps from this encoding along with some coordinates to effective deflection angles at some redshift. Finally, the components and intended workflow of the generation pipeline are presented in section 5.3.4. All networks were implemented using modules from `pyTorch` (Paszke *et al.*, 2019).

### 5.3.1. Set Model

A common approach for modeling strong gravitational lenses is to ascribe a parametric function to the mass profile of a lens galaxy and its dark matter halo. Such a function can generally be described by a few *lens parameters*, and can be related to the galaxy’s lensing effect through deflection angles computed from eq. (5.2.2).

In the context of multi-plane lensing, one would generally wish to model an arbitrary number of LOS perturbers at different redshifts in a given comoving volume. Provided that all these halos are described by the same type of functional mass profile, a population of halos can be represented as a *set* of equal-dimension vectors, where a given vector holds an individual halo’s lens parameters, position and redshift, making the order of the vectors of no importance. From a machine learning perspective, this calls for a permutation invariant model capable of processing a list of unknown length to extract its information to an encoding of fixed dimension. Most importantly, the length of this encoding should be chosen so to induce data compression, as the model can learn to represent global information about the set and disregard elements that have little impact on the objective. This aspect already gives the machine learning approach an edge in efficiency over true MPRT, which will involve equal computations for each individual halo, no matter how negligible its effect may be in the deflection angles.

Moreover, the paths of rays travelling through multiple planes of deflections are determined by a non-linear coupling between planes, as highlighted by the recursion in eq. (5.2.6). Therefore, the effect of different halos on effective deflection angles depends not only on their mass, but their relative positions in three-dimensional comoving space. In addition to being able to process sets,

the encoding model should then also benefit from the capacity to capture pairwise and higher-order interactions between halos in the set. Such criteria are met by a class of permutation invariant, attention-based neural networks known as Set Transformers (Lee *et al.*, 2019).

A mainstay in the field of natural language processing, the Attention mechanism is particularly well-suited for sequential data. For example, applying Attention between an input sentence and itself (known as self-attention) allows to capture long-range dependencies between words, which is crucial for text comprehension and language translation tasks. More formally, the Attention mechanism is a mapping from input query vectors to outputs using a set of key-value vector pairs, allowing to capture global dependencies between input and output. The Transformer architecture (Vaswani *et al.*, 2017) introduces Multi-head attention, where the query-key-value vectors  $Q, K, V$  are linearly projected multiple times to lower dimensional vectors, onto which Attention functions are independently applied and the outputs are concatenated. Using multiple attention heads allows the Transformer to capture more diverse interactions than single-head attention.

A Set Transformer makes use of attention-based set operations to apply multi-head self-attention to a set of inputs, and then aggregates a set of learned representations by an attention-based pooling operation. This casts the architecture in a typical encoder-decoder structure, where stacks of Set Attention Blocks (SAB) encode high order interactions between set elements, a Pooling by Multi-head Attention layer (PMA) pools these to a single representation, and linear layers decode this to a fixed-size vector providing a summary of the input set.

For an input set of length  $n$ , the self-attention computed by SABs is of  $\mathcal{O}(n^2)$  complexity. This proves to be prohibitively expensive in the case of large sets, as can be the case for our large halo populations. To this effect, Lee *et al.* (2019) define the Induced Set Attention Block (ISAB), where  $m$  learnable inducing points  $I$  are defined as part of the ISAB, with hyperparameter  $m < n$ . Multi-head attention is applied between these  $I$  and the input set rather than direct self-attention, improving the time complexity to  $\mathcal{O}(mn)$ .

From these components, the architecture of our Set Model can be described as follows

$$s_{\vartheta}(h) = \text{FC}(\text{PMA}(\text{ISAB}_m(\text{ISAB}_m(h)))) \tag{5.3.1}$$

where  $h$  is an input set of halos, FC is a fully-connected linear layer and  $\vartheta$  are the parameters of the Set Model. This model serves to compress the information of  $h$ , i.e. the parameters and positions of a variable-size population of halos, to a fixed size encoding.

### 5.3.2. Map Model

The functional role of the Map Model is to learn an  $\mathbb{R}^2 \rightarrow \mathbb{R}^2$  mapping from observed angular coordinates on the sky to the deflection angles, given a set of halos represented by an encoding provided by a Set Model. This is simply achieved by a fully connected multi-layer perceptron (MLP) that receives as input angular coordinates  $\boldsymbol{\theta}$ , redshift  $z$  at which to predict angles and the halo population encoding  $s_{\vartheta}(h)$ . Its output is a vector of effective deflection angles  $\boldsymbol{\alpha}^{(z)}$ .

Most use cases of a ray-tracer for strong lensing call for computing deflection angles on a pixelated grid. In light of this, we design the Map Model as a convolutional neural network (CNN) with  $1 \times 1$  convolutions, which is analogous to an MLP where the number of neurons per layer corresponds to the number of kernels per convolutional block. Such  $1 \times 1$  convolutions forgo any spatial correlation in the feature maps, and make the Map Model a pixel-wise function compatible with grids of any size. This greatly simplifies batching over pixels. A single input to this CNN then begins with a grid of angular coordinates  $\boldsymbol{\theta}$ , with the horizontal and vertical components split between two channels. The Set Model’s encoding  $s_{\vartheta}(h)$  and conditioned redshift  $z$  are split across additional channels and repeated over all pixels.

### 5.3.3. Objective function

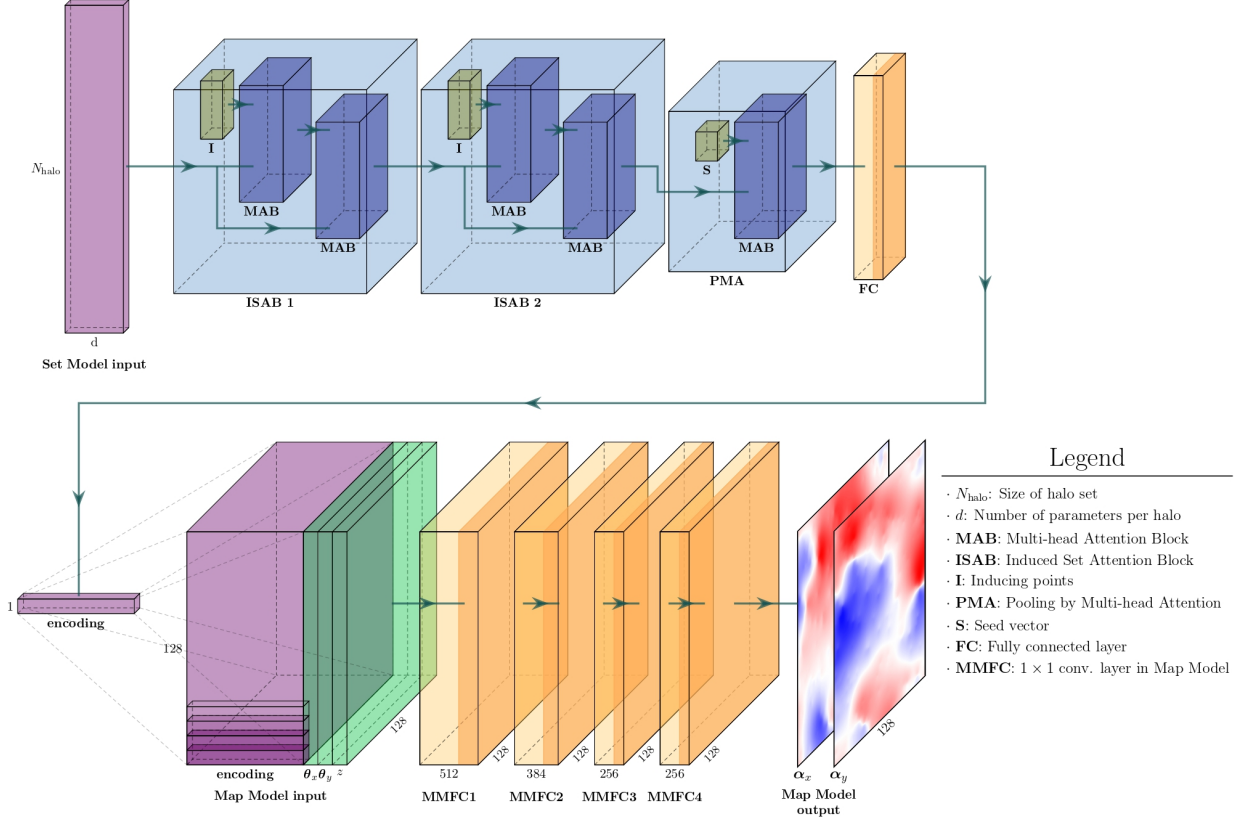
Given a set of halo parameters  $h$ , angular coordinates  $\boldsymbol{\theta}$  (and optionally outgoing ray directions  $\hat{\boldsymbol{\alpha}}_{\ell}$ ) and ground truth effective deflection angles  $\boldsymbol{\alpha}^{(z)}$  at redshift  $z$ , we aim to minimize the mean squared error over pixels,

$$\mathcal{L}_{\varphi, \vartheta}(h, \boldsymbol{\theta}, \boldsymbol{\alpha}^{(z)}) = \frac{1}{M} \sum_{j=1}^M \left( m_{\varphi}(s_{\vartheta}(h), \boldsymbol{\theta}_j, z) - \boldsymbol{\alpha}_j^{(z)} \right)^2, \quad (5.3.2)$$

where the index  $j$  labels corresponding pixels in the input and output,  $M$  is the total number of pixels and  $m_{\varphi}$ ,  $s_{\vartheta}$  are neural networks representing the Map and Set models, respectively.

### 5.3.4. Generation pipeline

The main design requirement for our pipeline is that it should only provide the deflection contribution from LOS structure, while having the freedom to choose any lens model for the main deflector plane. However, from the recursive nature of eqs. (5.2.4) and (5.2.6) for MPRT, any deflections in the background of the main deflector plane depend on all previous deflections. This implies that the pipeline must be broken up into two parts: one for the foreground and another for the background, while leaving it up to the user to define a main deflector model in the middle.



**Fig. 5.2.** Diagram of our network architecture for the Foreground Model. The network shown on top is a Set Model, which operates layers of self-attention and pooling on an input set of halo parameters. On the bottom is a Map Model, receiving an encoding of the halo population along with initial ray angles at  $z_0$  and a conditioned redshift  $z$  to produce effective deflection angles at said redshift. Shown below the  $1 \times 1$  convolution blocks of the Map Model are the number of filters, and along their depth is the side dimension of the square feature maps. The Background Model’s architecture is identical to this except for two additional channels in the Map Model input for the outgoing ray angles  $\hat{\alpha}_\ell$  at  $z_\ell$ .

Thus, the AlphaGen pipeline is composed of two independently trained, albeit near-identical models. The Foreground Model produces effective deflection angles at  $z_\ell$  conditioned on a population of halos in the foreground of the main deflector plane. The Background Model, however, produces a vector field of "angles"  $\Delta\alpha_{\text{bg}}$  that represent the contribution of the background halo population to the total effective deflections at  $z_s$ , removed from the deflections of the foreground and main deflector. This quantity is defined as follows:

$$\begin{aligned}
\beta &= \theta - \alpha_{\text{eff}}(\theta) \\
&= \theta - (\alpha_{\text{eff}}(\theta) - \alpha_{\text{eff, no bg}}(\theta)) - \alpha_{\text{eff, no bg}}(\theta) \\
&= \theta - \Delta\alpha_{\text{bg}}(\theta) - \alpha_{\text{eff, no bg}}(\theta).
\end{aligned} \tag{5.3.3}$$

In the above,  $\alpha_{\text{eff, no bg}}$  are effective deflection angles at  $z_s$ , but having been ray-traced through a background volume devoid of halos. These thus still contain the contribution of the foreground population as well as the main deflector plane, but not of the background halos. By itself, the quantity  $\Delta\alpha_{\text{bg}}$  has no physical meaning, as it is implicitly defined from previous deflections and not equivalent to raytracing through a volume that only contains the background<sup>1</sup> halos. The intent of this scheme is to design a Background Model that only targets the deflection signal of a background halo population, as this signal would otherwise be buried under the much stronger deflections of the main deflector model. The Background Model is conditioned on a population of halos in the background, as well as the outgoing ray angles and deflected coordinates after the main deflector plane. These two quantities depend on the foreground deflections as well as those of a user-defined main deflector. Both the Foreground and Background models are composed of a Set Model coupled to a Map Model. The architecture of one such network is illustrated in Figure 5.2.

As the purpose of our pipeline is to act as an accelerated stand-in for full MPRT, its intended workflow is designed somewhat analogously to the multi-plane lensing equation. Drawing from the form of eq. (5.2.6), the role of the Foreground Model comes in to obtain the angular coordinates at  $z_\ell$ :

$$\begin{aligned}\boldsymbol{\theta}_\ell &= \boldsymbol{\theta} - \sum_{i=1}^{\ell-1} \frac{D_{i\ell}}{D_\ell} \hat{\boldsymbol{\alpha}}_i(\boldsymbol{\theta}_i) \\ &= \boldsymbol{\theta} - F(h_F, \boldsymbol{\theta}, z_\ell)\end{aligned}\tag{5.3.4}$$

and the Background Model, drawing from eq. (5.2.4), appears as part of the expression to obtain the angular coordinates at  $z_s$ :

$$\begin{aligned}\boldsymbol{\beta} &= \boldsymbol{\theta} - \sum_{i=1}^N \frac{D_{is}}{D_s} \hat{\boldsymbol{\alpha}}_i(\boldsymbol{\theta}_i) \\ &= \boldsymbol{\theta} - \Delta\alpha_{\text{bg}}(\boldsymbol{\theta}) - \alpha_{\text{eff, no bg}}(\boldsymbol{\theta}_\ell, \hat{\boldsymbol{\alpha}}_\ell) \\ &= \boldsymbol{\theta} - B(h_B, \boldsymbol{\theta}_\ell, \hat{\boldsymbol{\alpha}}_\ell, z_s) - \alpha_{\text{eff, no bg}}(\boldsymbol{\theta}_\ell, \hat{\boldsymbol{\alpha}}_\ell)\end{aligned}\tag{5.3.5}$$

where  $F$  and  $B$  identify the Foreground and Background models, with  $h_F$  and  $h_B$  their respective halo sets, and  $\hat{\boldsymbol{\alpha}}_\ell$  are the physical deflection angles of the main deflector at  $z_\ell$ . The generation pipeline is then simply laid out with the following steps:

- (1) Sample a population  $h$  of LOS halos from a halo mass function

---

<sup>1</sup>"Background" only refers to the redshift interval of the volume here, as it loses its sense if there is no main deflector plane.

- (2) Forward pass the foreground set  $h_F = h(z < z_\ell)$  and the image plane coordinates  $\boldsymbol{\theta}$  through  $F$
- (3) Evaluate a main deflector model  $\hat{\boldsymbol{\alpha}}_\ell$  on the main deflector plane coordinates  $\boldsymbol{\theta}_\ell$ , obtained through eq. (5.3.4)
- (4) Forward pass the background set  $h_B = h(z > z_\ell)$ , the outgoing angles  $\hat{\boldsymbol{\alpha}}_\ell$  and coordinates  $\boldsymbol{\theta}_\ell$  through  $B$ .
- (5) Raytrace from the main deflector plane through an empty background volume, starting from coordinates  $\boldsymbol{\theta}_\ell$  and outgoing angles  $\hat{\boldsymbol{\alpha}}_\ell$ . This single-plane operation allows to obtain  $\boldsymbol{\alpha}_{\text{eff, no bg}}$
- (6) Effective deflection angles at the source plane are produced through  $\boldsymbol{\alpha}_{\text{eff}} = B + \boldsymbol{\alpha}_{\text{eff, no bg}}$

This simple pipeline then yields effective deflection angles at the source plane through two evaluations of GPU-accelerated neural networks, bypassing an expensive recursion over multiple interloping lens planes, while accurately simulating the LOS structures' lensing effect.

## 5.4. Data

This section describes the necessary components for simulating multi-plane deflection angles by ray-tracing through populations of line-of-sight halos. We first describe the analytic mass profile of individual halos in section 5.4.1, the sampling of line-of-sight halo populations in section 5.4.2, and finally the ray-tracing procedure to produce the ground truth effective deflections in section 5.4.3.

### 5.4.1. Halo mass profile

We model the radial density profile of individual halos with NFW profiles (Navarro *et al.*, 1997)

$$\rho(R) = \frac{\rho_0}{(R/R_s)(1 + R/R_s)^2} \quad (5.4.1)$$

where  $\rho_0$  is a characteristic density in units of  $M_\odot/\text{Mpc}^3$ ,  $R$  is the radial distance from the halo center in units of Mpc and  $R_s$  is a scale radius in units of Mpc. Instead of parameterizing our NFW profiles with their mass and concentration, we choose to describe them with parameters in angular units with  $r_s$  and  $\alpha_s$ , respectively defined as:

$$r_s = \frac{R_s u}{D_i} \quad (5.4.2)$$

$$\alpha_s = \frac{4\rho_0 R_s^2 (1 - \log 2) u}{\Sigma_c D_i} \quad (5.4.3)$$

given here in arcseconds, with  $u$  a rad-arcsecond conversion factor,  $D_i$  the angular diameter distance to a NFW halo  $h^{(i)}$  and  $\Sigma_c$  the critical surface density as described in eq. (5.2.3). A given halo  $h^{(i)}$  is then represented in the set  $h$  by the row of values  $(r_s^{(i)}, \alpha_s^{(i)}, x_0^{(i)}, y_0^{(i)}, z^{(i)})$ , where  $x_0, y_0$  are the angular coordinates of the halo’s center. Computing the projection integral of the three-dimensional density profile yields the projected surface density – or convergence –  $\kappa(\boldsymbol{\theta})$ . From this quantity, the deflection angles are obtained through an analytic solution to the convolution of the convergence, as in equation (5.2.2). This then yields a functional form for NFW deflection angles that can be evaluated on any angular coordinate grid.

#### 5.4.2. Sampling halo populations

We use the open-source software `pyHalo` (Gilman *et al.*, 2021) to sample populations of LOS halos by drawing masses from a redshift-dependent halo mass function in a given comoving volume, and to relate these masses to NFW parameters using a mass-concentration relation. `pyHalo` draws LOS halos in discrete redshift bins of width  $\Delta z$  from a Sheth-Tormen mass function (Sheth *et al.*, 2001). Throughout this paper, we assume a flat  $\Lambda$ CDM cosmology, using parameters from the Planck 2018 results (Planck Collaboration *et al.*, 2020). Our comoving volume is described by a double-cone geometry, opening out from redshift  $z_o$  by some opening angle up to the main deflector plane at  $z_\ell$ , then closing in towards the source plane at  $z_s$ . The cone in the background of the main deflector plane does not close exactly at  $z_s$ , but further behind it. This allows halos to be sampled up to the source plane. For this work, we have fixed the main deflector redshift at  $z_\ell = 0.5$  and the source redshift at  $z_s = 1.5$ . We choose the cone’s opening angle to be larger than the span of the first image plane; this ensures that the volume traversed by the rays overlaps with the rendering volume containing LOS halos. This can be visualized as a square pixel grid of some angular size being completely contained within a circular cross section of the cone.

It must be noted that the angular diameter distances between lens planes are computed assuming a flat Friedmann-Lemaître-Robertson-Walker (FLRW) background cosmology, which itself supposes homogeneity. However, the sampling of a halo population will add mass to these lens planes, causing the double-cone geometry to be locally overdense with respect to the mean density of the FLRW universe. If not taken into account, this will introduce an artificial focusing of the light rays towards the center of the cone. `pyHalo` corrects for this by adding sheets of negative convergence along the LOS that compensate the overdensity and bring the average surface density at each plane to the background value of our flat cosmology, as proposed by Birrer *et al.* (2017b). The need for this procedure adds to the computational toll of full MPRT, and further motivates the

development of an accelerated approach. Indeed, if the training data of our network is composed of sets of halo parameters without negative convergence sheets and their associated deflection angles, themselves computed by full ray-tracing including these sheets’ corrections, the network will implicitly learn to produce deflection angles that conserve the mean background FLRW density just by being conditioned on an input halo population.

### 5.4.3. Producing datasets

The training sets for the Foreground model are constructed by sampling populations of LOS halos at  $z < z_\ell$  and ray-tracing through them up to  $z_\ell$ , starting from zero-deflection initial conditions at  $z_o$ . We used the open source software `lenstronomy` (Birrer et Amara, 2018; Birrer *et al.*, 2021) for the multi-plane lensing computations. For a given training example, the halos are saved as a list of NFW parameters and halo redshift, and ground truth effective deflection angles are saved at  $z_\ell$ .

As one of the design requirements of the pipeline is to leave the choice of main deflector up to the user, careful considerations must be taken when constructing training datasets for the Background Model. Since all background light ray trajectories depend on previous deflections from the foreground and the main deflector through the recursion in eq. (5.2.6), the nonzero-deflection initial conditions at  $z_\ell$  need to be defined as to reflect the possible realm of main deflectors that could be chosen in all use cases of the pipeline. In the  $\mathbb{R}^D$  (with  $D = 2n_{\text{pix}}$ ) space of outgoing ray angles after  $z_\ell$  that serves as part of the input to the Background model’s Map Model, a general approach is to choose pixel values randomly drawn from a uniform distribution in some symmetric and zero-centered range of plausible angle values. We can afford to employ this scheme because the Map Model has been defined as a pixel-wise function, therefore there need not be any spatial correlation in the input or output spaces of the Map Model. Choosing this training regime for the Background model makes it generalizable to any reasonable choice of main deflector. The smooth outgoing rays associated to the lensing from a realistic galaxy represent a manifold in  $\mathbb{R}^D$  that is exceedingly unlikely to have been sampled among the randomized initial conditions, but the individual pixel values will be sure to have been covered by the random sampling scheme, given a training dataset of sufficient size.

A single training example for the Background model is thus constructed by the following steps:

- (1) Sample a population of LOS halos in the full span of the geometry, i.e.  $z_o < z < z_s$
- (2) Ray-trace through the foreground halos up to  $z_\ell$



- (3) Obtain the deflected coordinates  $\boldsymbol{\theta}_\ell$  from the effective foreground deflections, and the outgoing ray angles  $\hat{\boldsymbol{\alpha}}_\ell$  from the randomized prescription outlined above
- (4) Use these coordinates and ray angles as starting points for ray-tracing through the background halos up to  $z_s$ , to obtain  $\boldsymbol{\alpha}_{\text{eff}}$
- (5) Use these same coordinates and ray angles as starting points for ray-tracing through an empty background up to  $z_s$ , to obtain  $\boldsymbol{\alpha}_{\text{eff, no bg}}$
- (6) Produce the training example’s label  $\Delta\boldsymbol{\alpha}_{\text{bg}} = \boldsymbol{\alpha}_{\text{eff}} - \boldsymbol{\alpha}_{\text{eff, no bg}}$

The full halo population is then saved as a list of NFW parameters and halo redshifts, along with the outgoing ray angles and deflected coordinates at  $z_\ell$ . The vector field  $\Delta\boldsymbol{\alpha}_{\text{bg}}$  is saved as the associated label.

Two independent datasets of 25 000 examples were produced at  $128 \times 128$  pixels for the Foreground and Background models respectively. For both datasets, LOS NFW halos were sampled from a Sheth-Tormen mass function in the  $10^7 - 10^{10} M_\odot$  interval within a fixed double-cone geometry at an opening angle of  $11.0''$ , then ray-tracing in a  $7.68''$  field of view. This corresponds to a  $0.06''$  pixel resolution, comparable to that of HST’s ACS instrument. Due to technical limitations preventing the batching of `pyTorch` tensors of variable lengths<sup>2</sup>, the halo sets were sampled at a fixed size  $\langle N_{\text{halo}} \rangle$  corresponding to the expected total number of halos in our comoving volume. The pixel values in units of arcseconds for the  $\hat{\boldsymbol{\alpha}}_\ell$  inputs of the Background Model were sampled from the distribution  $\mathcal{U}(-5, 5)$ . A standard 80/20 split was used for separating datasets into training/validation subsets.

## 5.5. Training

### 5.5.1. Feature normalization

In order to stabilize and facilitate the training of our networks, feature normalization was employed for both datasets. For any given dataset, we record in a file the minimum and maximum values across all examples for the input features, i.e. halo parameters and lensing quantities, and the absolute maximum values of the ground truth effective deflections at their respective redshifts. Feature-wise min-max scaling to the interval  $[-1, 1]$  using dataset bounds is applied to selected NFW parameters, namely  $(r_s, \alpha_s, x_0, y_0)$ , while leaving the halo redshifts unchanged as they are of order unity. For  $r_s$  and  $\alpha_s$ , this scaling is applied to the log of their values, as they are strictly positive and can vary across many orders of magnitude for different halos. The input lensing

<sup>2</sup>As of `pyTorch v1.11`, which was used for the implementation of this pipeline

quantities  $\theta$ ,  $\hat{\alpha}_\ell$  and  $\theta_\ell$  are also normalized by min-max scaling to  $[-1, 1]$  using dataset bounds. Lastly, the ground truth quantities  $\alpha_{\text{eff}}$  and  $\Delta\alpha_{\text{bg}}$  are divided by their absolute maximum values at their respective redshift, taken from the dataset. To avoid computing such normalizations at every batch during training, an entire dataset is loaded, normalized, then saved in a temporary file at the start of training. These transformations and their inverses are defined as part of a "model" object, so that any quantity of interest can be readily re-scaled for use cases of the pipeline.

### 5.5.2. Foreground Model

The architectures of both components of the Foreground Model were grid-searched to identify a set of relevant hyperparameters. In a Set Transformer of the form (5.3.1), Lee *et al.* (2019) recommend the use of a SAB after the aggregation of the PMA only when producing  $k > 1$  outputs, in order to capture interactions between elements. However, our Set model only outputs a single encoding of a halo set. In our experiments comparing the use of a SAB and FC layer in the decoder end of the Set Model to a FC layer only, we found the latter to lead to a lower value of the objective function (5.3.2) while being more efficient.

**Table 5.1.** Detailed architecture of the Foreground and Background models.

Set Model	Map Model
ISAB <sub>128</sub> (128, 8)	Conv2D(512, 1, 1, ELU)
ISAB <sub>128</sub> (128, 8)	Conv2D(384, 1, 1, ELU)
PMA <sub>1</sub> (128, 8)	Conv2D(256, 1, 1, ELU)
FC(256, -)	Conv2D(256, 1, 1, ELU)

**Table 5.2.** Training hyperparameters of the Foreground and Background models

Parameter	Value
Optimizer	Adamax
Learning rate	$10^{-3}$
Batch size	64 ( <i>F</i> ) / 32 ( <i>B</i> )

We report the best performing architectures for the Foreground Model’s Set and Map models in Table 5.1, and other hyperparameters in Table 5.2. The terminology for the layers listed in Table 5.1 is as follows: in the Set Model, ISAB<sub>*m*</sub>(*d*, *h*) denotes an ISAB with *d* units, *h* heads and *m* inducing points, PMA<sub>*k*</sub>(*d*, *h*) denotes a PMA with *d* units, *h* heads and *k* seed vectors, FC(*d*, *f*) denotes a FC layer with *d* units and activation function *f*. In the Map Model, Conv2D(*d*, *k*, *s*, *f*) is a 2D convolutional layer with *d* output channels, kernel size *k*, stride *s* and activation function *f*. We trained this model for 6000 epochs, amounting to approximately 9 days wall time in batch parallel on 4 Nvidia V100 GPUs.

### 5.5.3. Background Model

A less extensive grid-search was conducted for the Background Model’s architecture and hyperparameters, as the best configuration found for the Foreground Model proved to be effective here as well. We trained the Background Model for 8500 epochs, amounting to approximately 22 days wall time in batch parallel on 4 Nvidia V100 GPUs. The larger halo sets in the background segment of the lensing system account for the longer time per epoch.

## 5.6. Results and Discussion

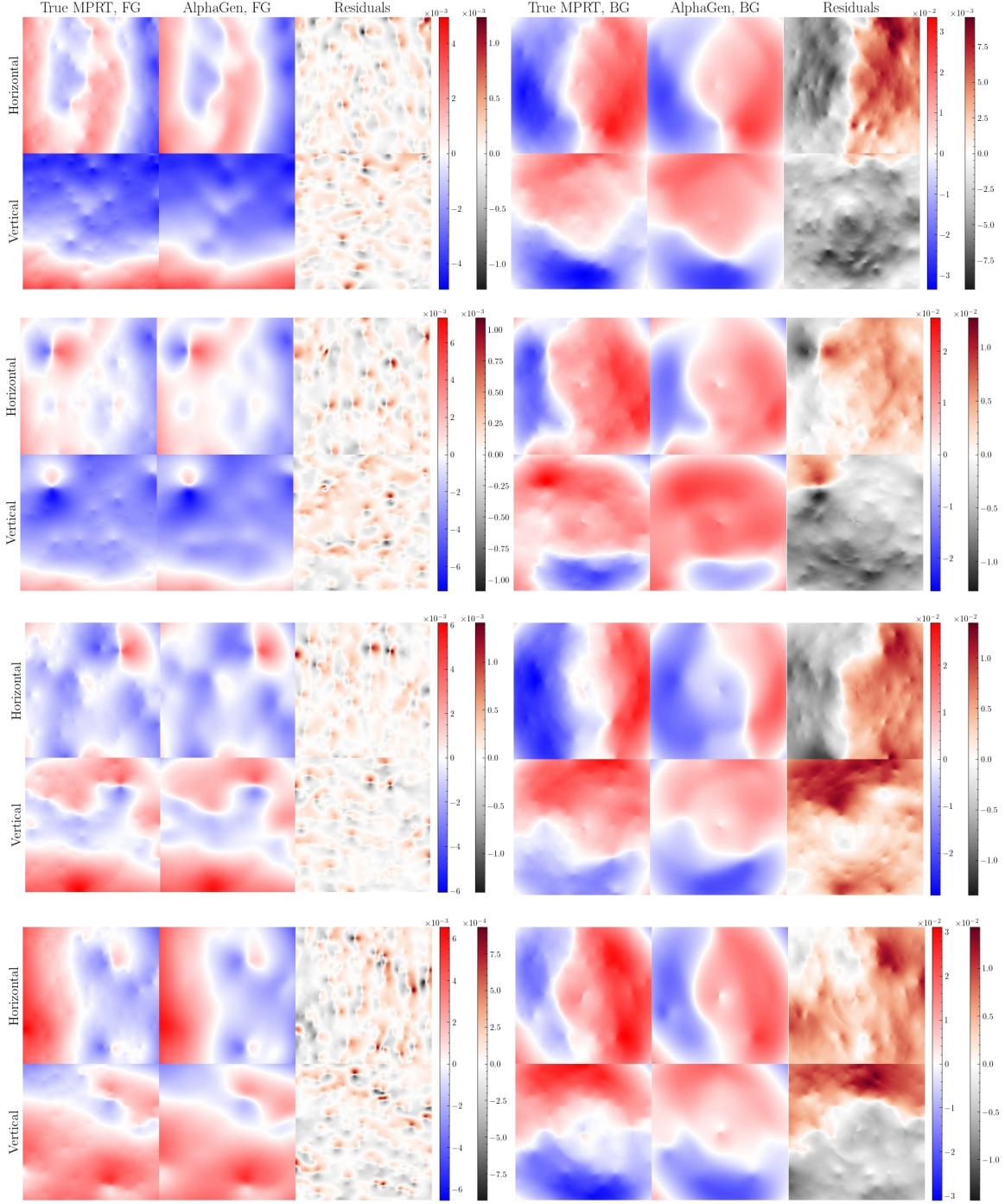
In this section, we present the performance of AlphaGen in various aspects. In section 5.6.1, we discuss the performance of our pipeline on randomly sampled test examples of halo populations never seen by the network in training. Section 5.6.2 discusses how our pipeline compares to true MPRT through the lensing of source galaxies. Finally, section 5.6.3 highlights the speed gain from our pipeline relative to true MPRT.

**Table 5.3.** Main deflector model parameters.

Parameter	Distribution
EPL	
Halo mass $\log_{10}(M/M_{\odot})$	$\mathcal{U}(11.7, 12.0)$
Negative power-law slope $\gamma$	$\mathcal{U}(1.8, 2.0)$
Complex ellipticity moduli $e_1, e_2$	$\mathcal{U}(-0.2, 0.2)$
Halo position $x_0, y_0$ (")	$\mathcal{U}(-0.1, 0.1)$
Shear	
Shear components $\gamma_1, \gamma_2$	$\mathcal{U}(-0.2, 0.2)$

### 5.6.1. Test set performance

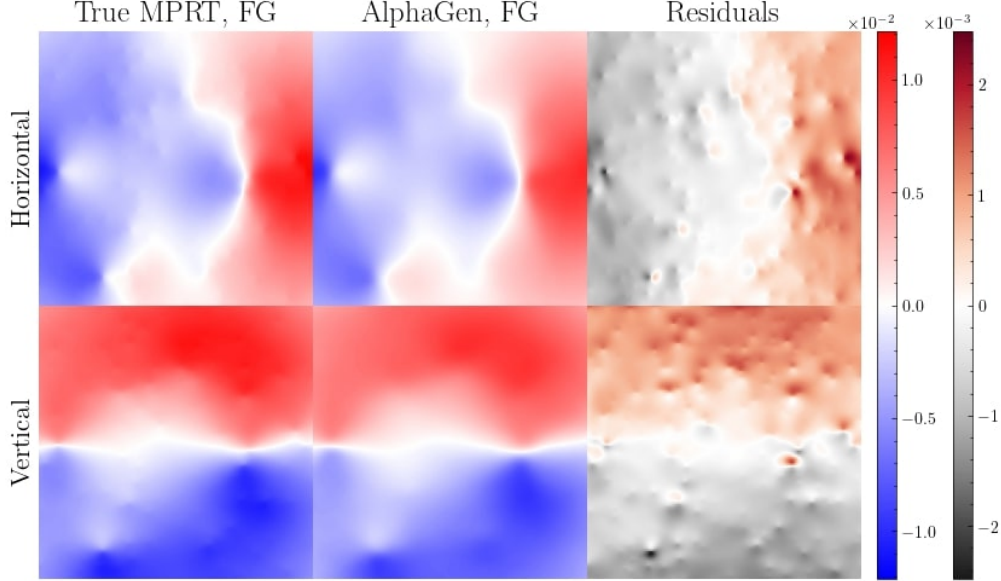
Figure 5.3 shows some sample reconstructions of foreground effective deflection angles and background halo contributions to total deflection angles, respectively produced by Foreground and Background Models. We model the main deflectors as elliptical power law (EPL) mass profiles (Tessore et Metcalf, 2015) with added external shear, with parameters drawn from distributions listed in Table 5.3. All deflector models parameters sampled from the distributions in Table 5.3 for these results led to values of outgoing angles  $\hat{\alpha}_{\ell}$  contained within the bounds chosen for the Background Model’s training set. Furthermore, the test halo sets were sampled from the same mass function as well as comoving volume,  $z_{\ell}$  and  $z_s$  as for the training sets described in section 5.4.3. Without sacrificing generality, these tests do not model subhalo populations in the main



**Fig. 5.3.** Foreground (left column) effective deflection angles  $\alpha_{\text{eff}}$  and background (right column) halo contribution  $\Delta\alpha_{\text{bg}}$  to total deflection angles associated to four test halo populations and main deflector models (along rows), comparing angles computed from true MPRT and from the Foreground and Background models of the AlphaGen pipeline. The background angles are typically of a higher amplitude than those of the foreground, due to the larger comoving volume and number of halos. Reconstructions of angles at a pixel size of  $128 \times 128$ , corresponding to the size of these models' training labels.

deflector plane, as the pipeline is compatible with any main deflector model. We observe in the left column of Figure 5.3 that the predicted foreground angles are in excellent agreement with the true angles at large scales, and that there is good correspondence between the positions of halos responsible for the largest perturbations in the deflection field. However, the general smoothness of the predicted angles as well as the high-frequency residuals show that the model struggles to capture the smallest scale perturbations, symptomatic of the known *spectral bias* (Rahaman *et al.*, 2018) of neural networks. This behavior is to be expected from the data compression induced by the Set Model. As it learns to summarize a large halo set to an encoding that will minimize the loss function, compression necessarily implies that the halos with the least impact on the effective deflection angles, be it because of their small mass or unfavorable relative position in  $z$ , might be overlooked in the encoding. Furthermore, we observe that the mean absolute scale of the residuals is generally around 2% that of the ground truth angles. The worse pixels in the residuals often correspond to the positions of some smaller deflections from halos that were missed by the model, with the absolute maximum of the residuals generally being around 20% that of the ground truth angles.

The background halo effective deflection angle contributions output by the Background Model are compared to the true  $\Delta\alpha_{\text{bg}}$  values in the right column of Figure 5.3. Our model can generally reproduce the true  $\Delta\alpha_{\text{bg}}$  accurately at large scales, and we can observe some correspondence between the positions of individual deflectors. Nonetheless, the residuals indicate a poorer performance than the Foreground Model. This is to be expected, since the diversity of its input features make the task of the Background Model a more difficult problem. Ray-tracing in the foreground starts from zero-deflection initial conditions, whereas the background ray-tracing depends on coordinates deflected by the foreground as well as deflections from the main lens. In the context of this neural network, this implies that the gradient flow from the objective function is more evenly divided between the  $(\theta_\ell, \hat{\alpha}_\ell)$  inputs of the Map Model and the halo encoding from the Set Model, while the unchanging  $\theta$  input of the Foreground Model’s Map Model carried much less weight in its predicted values. In the case of the Background Model, the mean absolute residual scale is generally around 20% that of the ground truth angles, with the absolute maximum of the residuals typically being more than 50% that of the ground truth values. The complex input feature space of the Background Model could warrant training with a larger dataset in order to improve performance, or devising an alternative scheme for the production of training examples instead of the one presented in section 5.4.3.



**Fig. 5.4.** Example of the reconstruction of foreground angles associated to a test halo population of a different size than the training halo populations. While the intricate spatial distribution of the angles is well reproduced, there is a manifest directional bias in the residuals.

The test halo populations associated to the reconstructions shown in Figure 5.3 were sampled at the same fixed set size  $\langle N_{\text{halo}} \rangle$  as the training sets, which is the expected number of LOS halos for this geometry. Even so, our models are capable of processing halo sets of sizes they have not been trained on. Figure 5.4 displays one such example, where the test set size  $N_{\text{halo}}$  is drawn from a Poisson distribution with mean  $\langle N_{\text{halo}} \rangle$ . Although the predictions capture the general features of the angles well, there appears to be a coordinate-aligned bias in the difference between true and predicted angles. These residuals indicate that the predicted angles fall off in amplitude relative to the true angles along the coordinate directions. Although we observe this behavior for the particular example of Figure 5.4, it is not systematic, as we have observed the opposite to occur (i.e. a relative directional increase in amplitude) for other generations of test halo sets. We have found neither correlation or anticorrelation between  $N_{\text{halo}} > \langle N_{\text{halo}} \rangle$  and a directional increase in amplitude, or vice-versa for  $N_{\text{halo}} < \langle N_{\text{halo}} \rangle$ . This bias likely stems from the PMA block’s pooling operation within the Set Model as described in (5.3.1), whose weights were consistently trained to extract an encoding from internal representation sets of identical size. We believe that this bias could be readily addressed by training the models on halo sets of varying size.

### 5.6.2. Lensing performance

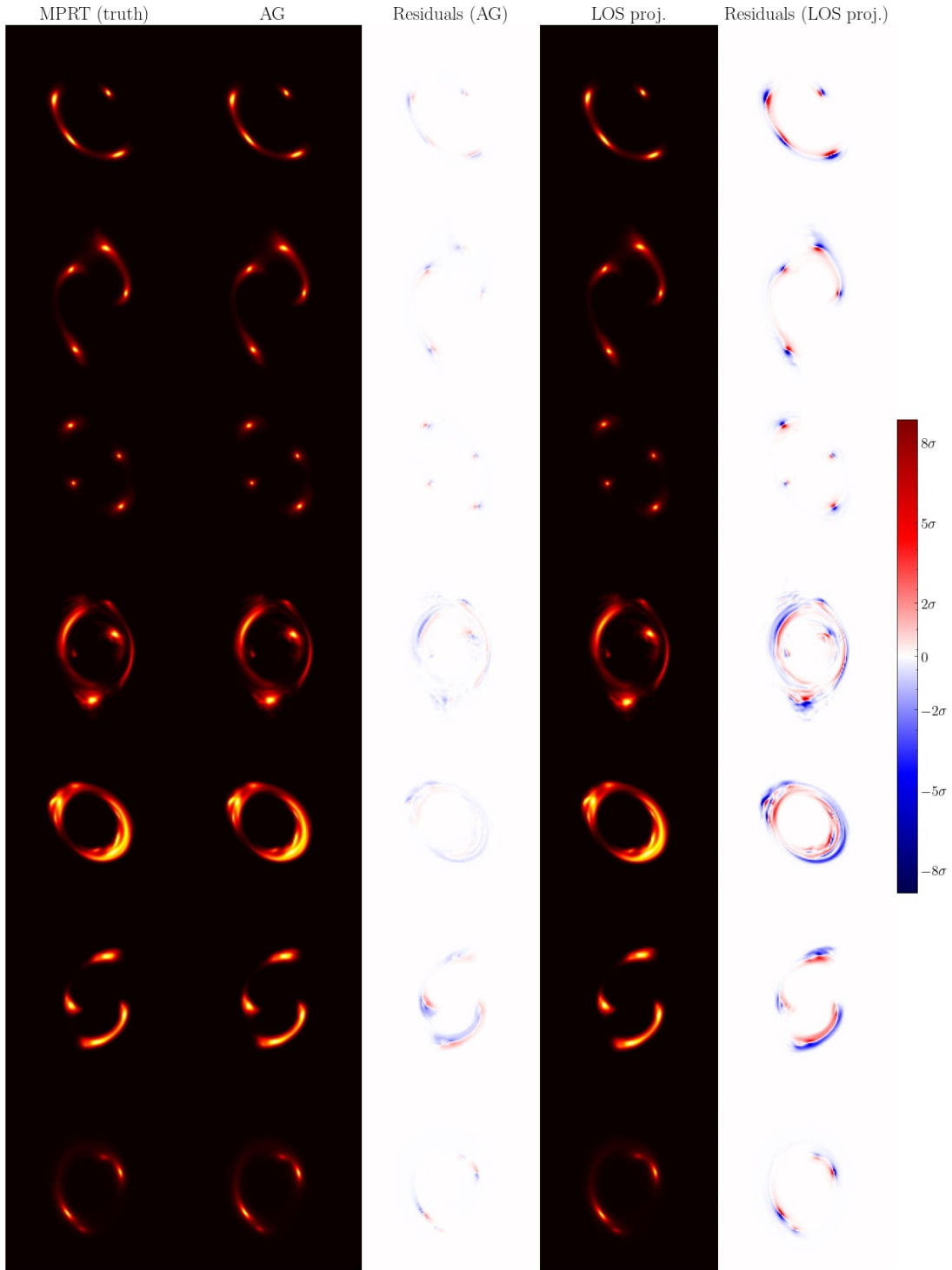
As the definitive test of AlphaGen’s performance is in its application to reproducing observables, we use the pipeline to model observations of gravitational lenses including LOS halos. For

the background source brightness distributions, we use processed galaxy images from the Cosmic Evolution Survey (COSMOS; Scoville *et al.*, 2007) taken by the Hubble Space Telescope’s (HST) Advanced Camera for Surveys (ACS). The images have been preprocessed to  $128 \times 128$  pixel postage stamps, normalized to pixel intensities in the range  $[0, 1]$  and denoised according to the procedure described in Adam *et al.* (2023). We choose a 4” field of view for the source plane, and otherwise use the same geometry and mass function parameters as described in section 5.4.3. We also model main deflectors with the same mass profile as in section 5.6.1, with parameters sampled from the distributions of Table 5.3.

As an additional point of comparison, we also model LOS halos following the prescription from Şengül *et al.* (2020). In this approximation for multi-plane lensing, LOS halos are treated as effective subhalos by modifying the scale radius and mass parameters of their density profiles. This projection of thousands of interloping halos to the main lens plane reduces the system to a single-plane lens, making for effective computations, but doing away with the non-linear effects that arise in MPRT. We also use these models to produce mock gravitational lenses of COSMOS background sources, referring to this modeling approach as the *LOS projection* scheme.

Figure 5.5 shows samples of these simulated lenses, comparing those produced by MPRT to those by the AlphaGen pipeline and by LOS projection. The MPRT, AlphaGen and LOS projection lenses share identical LOS halo populations and main deflector models, meaning that they only differ in their simulation of LOS effects. In galaxy-galaxy strong lenses, the contributions of LOS halo to the deflections are of the order of tens of milliarcseconds; they are thus buried under the deflections from the main lens, which are at the arcsecond level. Hence, the lenses from the different modeling schemes can hardly be distinguished by eye. In order to quantify the accuracy of the AlphaGen modeling pipeline, we take the residuals between the MPRT and AlphaGen lenses and scale them by some assumed noise level  $\sigma$ . Collett (2015) defines a signal-to-noise ratio (SNR) of at least 20 to be a required detectability criteria for finding strong lenses in upcoming optical imaging surveys. As such, we choose  $\sigma = 0.05$ , so that the peak SNR of our lensed galaxies in the  $[0, 1]$  range would be  $\text{pSNR} \sim 1/\sigma = 20$ . We follow the same procedure when taking the residuals between the MPRT and LOS projection lenses, to provide a reference in accuracy from this multi-plane lensing approximation.

The AlphaGen residuals in Figure 5.5 indicate that the errors introduced by AlphaGen are mostly localized on and around the multiple source images, as is to be expected. Some shifts of a few pixels in the deflections can displace the brightest regions of the observed lens and lead to  $\sim 2\sigma$  errors in the residuals, as can be observed in the case of some lenses. For problems where the target



**Fig. 5.5.** Simulated lensing of COSMOS sources of normalized intensity including deflections from LOS halos, with effective multi-plane deflection angles computed with true MPRT (first column) and the AlphaGen pipeline (second column). The residuals between these two models (third column) are expressed in terms of a noise level  $\sigma = 0.05$ , so that the pSNR the observations would be  $\sim 1/\sigma$ . Also shown are lensed COSMOS sources modeled with the LOS projection scheme (fourth column) and their associated residuals (fifth column) with respect to the true MPRT lenses.



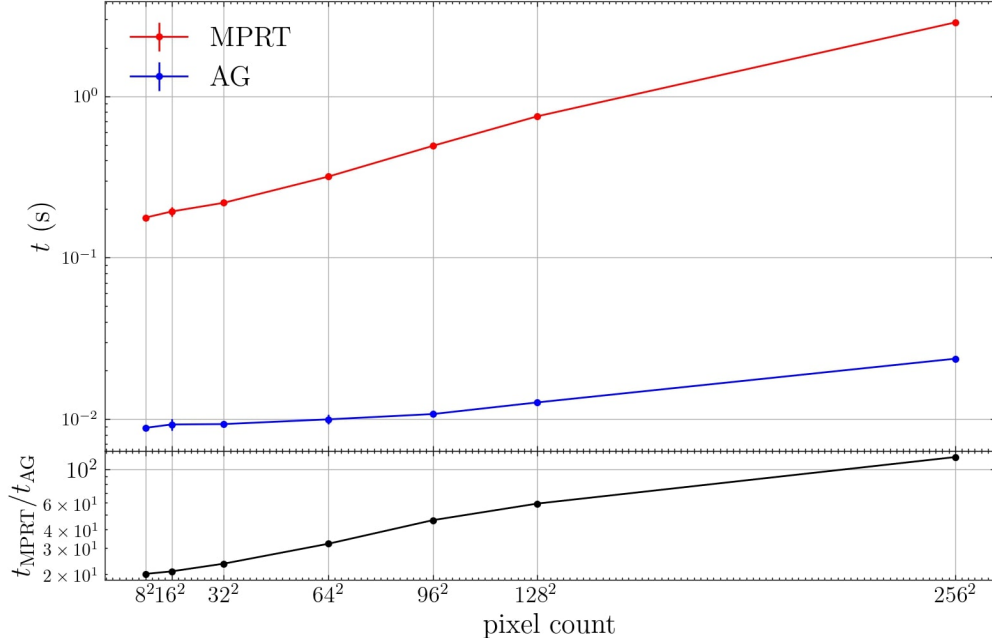
signal can be at the level of individual pixels, such errors are likely to lead to biased inference. However, Figure 5.5 also indicates that the errors incurred by the LOS projection method are much more significant and systematic across all examples than that from AlphaGen, often exceeding  $8\sigma$ . This could stem from the absence of non-linear lensing effects in this single-plane approximation to gravitational lensing from LOS halos. Comparing the residuals in the third and fifth columns of Figure 5.5 demonstrates that AlphaGen is more faithful to true MPRT than the LOS projection approximation.

Nevertheless, taking these shortcomings in accuracy into consideration, the current applicability of AlphaGen as a simulator for LOS structure inference problems is somewhat limited, especially for medium-to-high-SNR observations. As shown by the simulated deflection angles of Figure 5.3, AlphaGen’s modeling error is largely dominated by the Background Model inaccuracy, with errors at the scale of the lensing signal itself. Since the background volume of a double-cone lensing geometry is typically greater than the foreground’s, more LOS halos will be sampled in the background on average and lead to stronger deflections. This implies that the Background Model is the most important component for the accurate reproduction of MPRT deflection angles. Therefore, further improvements to this model’s performance must be made before the pipeline can be applied to such use cases.

### 5.6.3. Acceleration

Figure 5.6 compares runtimes for the generation of multi-plane deflection angles between our pipeline and the true MPRT formalism described in section 5.2.2. As our pipeline is composed of pixel-wise neural networks, it is compatible with grids of any number of pixels, allowing to test how its runtime scales when varying the pixel count. Changing the pixel resolution in this fashion implies that the Foreground Model might receive as input angular coordinates it has not seen in training. Yet, we observe it to interpolate adequately between training coordinates, be it for producing angles at a pixel size lesser or greater than the training size of  $128 \times 128$ .

The benchmarks in Figure 5.6 indicate that our pipeline accelerates the generation of multi-plane deflection angles by 1 to 2 orders of magnitude relative to the MPRT formalism, over the range of pixel counts tested. Additionally, the speedup scales up with pixel size, with our pipeline producing angles over 100 times faster at a size of  $256 \times 256$ . It is also important to note that the AlphaGen benchmarks shown in Figure 5.6 were calculated for a single model input at a time. Leveraging the efficiency of batching together multiple halo sets to simultaneously simulate their deflection angles could drastically increase the speedup provided by AlphaGen. These significant



**Fig. 5.6.** (Top) Runtime of the generation of multi-plane deflection angles as a function of the number of pixels in the field of view, for the true MPRT computation (red) and the forward passes of our pipeline’s networks (blue). For each pixel size, 10 generations of test halo populations of variable size were sampled and the production of their deflection angles was timed for both methods. The points above mark the mean runtimes across these 10 generations, and barely visible errorbars correspond to the standard deviation of the runtimes. Note that the time axis is in log-scale. (Bottom) Ratio of the mean runtimes between true MPRT and our pipeline.

advantages in efficiency illustrate the usefulness of our pipeline as a stand-in for true MPRT. Nevertheless, as discussed in section 5.6.2, AlphaGen’s accuracy is still not up to par for the exceedingly strict error budget required for unbiased inference of LOS structure in strong lenses.

## 5.7. Summary and Conclusion

In this work, we have presented AlphaGen, a neural network-accelerated simulation pipeline for the deflections of line-of-sight halos in strong gravitational lenses. Our method bypasses the main efficiency drawbacks of the traditional multi-plane ray-tracing framework by minimizing the number of recursions with a two-step approach. One model produces deflection angles for all halos in the foreground of a main lens plane simultaneously, and a separate model can be conditioned on these deflections as well as those of a main deflector to similarly produce the deflections of all background halos at once. This efficient pipeline is flexible, leaving the choice of mass profiles in the lens plane (main lens and subhalos) entirely up to the user.

Furthermore, the design of our neural network components as set-based models is of high value to small-scale structure modeling. AlphaGen is compatible with random populations of halos sampled from a halo mass function, whose sizes vary across different generations. Moreover, only CDM halo population were considered in this work, but one could construct a training dataset containing populations of halos associated to multiple different dark matter models, such as WDM, for which the number of low-mass halos is drastically suppressed. AlphaGen’s Map Models are also fully conditioned on the lensing geometry, implying that the pipeline would be compatible with more diverse geometries than the fixed configuration chosen for this work. With such expansions of the training domain, along with its compatibility with any pixel grid size, AlphaGen could become a general-purpose stand in for multi-plane ray-tracing.

This being said, the AlphaGen pipeline is still currently limited by its accuracy in the reproduction of deflection angles from multi-plane ray-tracing. Our tests in section 5.6 demonstrate that the Background Model introduces significant errors that could lead to biased inference of LOS structure in strong lenses, which is particularly sensitive to pixel-level brightness perturbations. We believe that this lackluster performance could be remedied by devising an alternative training scheme for the Background Model, or revising the manner of training label production altogether. We plan to explore such avenues in future works, as the development of efficient and accurate simulators of LOS structure effects in strong lensing for Bayesian inference problems is paramount for placing constraints on dark matter models.

## Software

The AlphaGen pipeline is still in active development, but its source code as well as various associated scripts and trained network parameters will eventually be made available as a standalone open-source software, and will be implemented within the upcoming open-source strong gravitational lensing package `caustics`. This research made use of `pyTorch` (Paszke *et al.*, 2019), `pyHalo` (Gilman *et al.*, 2021), `lenstronomy` (Birrer et Amara, 2018; Birrer *et al.*, 2021), `NumPy` (Harris *et al.*, 2020), `Matplotlib` (Hunter, 2007), `h5py` (Collette *et al.*, 2023), `TensorBoard` (Abadi *et al.*, 2015), `colossus` (Diemer, 2018) and `Astropy` (Astropy Collaboration *et al.*, 2013, 2018, 2022).

## Acknowledgements

We would like to thank Alexandre Adam and Laurence Perreault-Levasseur for many insightful discussions and their valued input throughout the development of this project. We also thank Daniel Gilman and Kyunghyun Cho for their insights and helpful suggestions. This research was supported

by the Schmidt Futures Foundation. This work was also enabled in part by computational resources provided by Calcul Québec, Compute Canada and the Digital Research Alliance of Canada. C.W. was supported by an ESP excellence scholarship from Banque de Montréal.

# Chapitre 6

---

## Conclusion

Dans ce mémoire, nous avons exploré le développement d'un nouveau formalisme basé sur l'apprentissage profond pour l'accélération de simulations du lentillage gravitationnel par des halos de matière sombre en ligne de visée de lentilles galaxie-galaxie. Les réseaux de neurones qui composent notre pipeline de simulation ont été développés afin de satisfaire l'exigence de conception d'avoir la capacité de traitement d'entrées sous forme d'ensembles. Cela offre à notre formalisme la capacité de modéliser complètement les effets de lentillage d'une population variable de halos de matière sombre tirés d'une fonction de masse, connectant directement les propriétés d'un modèle particulière de matière sombre aux observables. De plus, nos modèles opèrent sur des grilles de pixels de taille arbitraire, peuvent être entièrement conditionnés sur la géométrie du système observateur-lentille-source, et sont compatibles avec n'importe quel modèle de masse dans le plan de lentille principal.

Nos expériences mettant à l'épreuve la performance de notre simulateur ont démontré qu'il permettait une accélération de presque deux ordres de grandeur p.r.a lentillage à plans multiples. La performance de notre Modèle d'Avant-plan (*Foreground Model*) sur des exemples tests de populations de halos est excellente dans sa reproduction des angles véritables provenant du lentillage à plans multiples, malgré qu'il échoue à reproduire les perturbations aux plus petites échelles provenant des halos légers. Des travaux futurs iraient à caractériser l'impact de ces erreurs sur la simulation d'images de lentilles gravitationnelles, et d'étudier le biais introduit en relation à des populations de halos tirés d'une fonction de masse de WDM, pour lesquelles l'abondance de halos légers est fortement supprimée.

Dans le cas du Modèle d'Arrière-plan (*Background Model*), sa performance sur des exemples tests laisse à désirer. Tel que discuté dans l'article, le problème d'apprentissage de l'arrière-plan est plus complexe, puisque le modèle doit apprendre à capturer les interactions subtiles entre la

population de halos et les angles sortant du plan de lentille principal afin de reproduire l'objectif correctement. Or, nous hypothétisons que les tâches de ce modèle et de celui de l'avant-plan pourrait être facilitées en révisant notre moyen de produire les angles de déflexions effectifs qui sont les objectifs de ces modèles.

Une inspection de la Figure 5.3 identifie une caractéristique systématique à travers tous les champs de déflexion calculés par lentillage à plans multiples. On y remarque la séparation directionnelle rouge-bleu, représentant un dipôle à grande échelle de valeurs positives à négatives. Nous croyons que cette caractéristique n'a pas de véritable provenance physique, et qu'elle soit symptomatique de la procédure de modélisation des halos et des calculs de lentillage à plans multiples. En effet, comme nous générons des halos dans un certain volume de l'Univers, cela viole le principe cosmologique d'homogénéité et d'isotropie en créant une surdensité de matière dans ce volume particulier. Certes, nous employons une procédure servant à équilibrer cette surdensité dans chaque plan de lentille afin d'y retrouver la densité moyenne de l'Univers, tel que décrite en section 5.4.2. Or, cette procédure ne compense pas (ou pas suffisamment) pour l'absence artificielle de halos hors du volume de génération, dont le potentiel gravitationnel servirait à annuler l'accumulation de déflexions observée près des extrémités des champs de vision carrés en Figure 5.3. Dans des travaux futurs, nous prévoyons développer un formalisme mathématique apportant des modifications au lentillage à plans multiples afin de corriger ces effets. De plus, nous croyons que le choix de modèle de halos de champ peut y jouer un rôle; les modèles de masse NFW employés dans cet ouvrage ont un profil radial de densité diminuant asymptotiquement vers zéro, mais ont tout de même une masse intégrée totale infinie. Il est alors possible qu'il y aille une accumulation artificielle de déflexions provenant de ce surplus de masse en s'éloignant du centre du volume. Cet effet pourrait être contourné en utilisant des profils avec masse intégrée finie, tels qu'un NFW tronqué (Baltz *et al.*, 2009). Ce dipôle non-physique dans les angles de déflexions représente une part importante de l'intensité des pixels de l'objectif, et vient certainement masquer des véritables déflexions de halos. Cela nous porte à croire que de se débarrasser de cet effet viendrait faciliter l'entraînement et améliorer la performance des modèles.

Ce travail vient porter une contribution importante au champ d'étude de la structure légère de matière sombre en ligne de visée des lentilles gravitationnelles. Il s'agit d'une branche émergente en cosmologie, et la littérature à ce sujet est encore très jeune. Avec la prochaine génération de télescopes qui verront première lumière dans cette décennie, l'explosion attendue du nombre de systèmes candidats de contenir la signature de ces halos légers va stimuler une multiplication des

publications scientifiques sur le sujet. Parallèlement à cela, cette nouvelle ère de l'intelligence artificielle entraîne le développement de modèles puissants qui trouvent des multitudes d'applications en astrophysique et stimulent un essor de nouvelles découvertes scientifiques. Il s'agit alors du moment opportun pour établir un cadre de modélisation rigoureux et efficace du lentillage en ligne de visée, préparant le terrain pour ce qui promet de devenir l'un des meilleurs moyens de contraindre la nature de la matière sombre. Mes objectifs de recherche sont maintenant d'amener notre formalisme à son plein potentiel, afin de pouvoir l'appliquer à des problèmes d'inférence visant à détecter des populations de halos perturbateurs dans des données de lentilles gravitationnelles.





## Références bibliographiques

---

- Martín ABADI, Ashish AGARWAL, Paul BARHAM, Eugene BREVDO, Zhifeng CHEN, Craig CITRO, Greg S. CORRADO, Andy DAVIS, Jeffrey DEAN, Matthieu DEVIN, Sanjay GHEMAWAT, Ian GOODFELLOW, Andrew HARP, Geoffrey IRVING, Michael ISARD, Yangqing JIA, Rafal JOZEFOWICZ, Lukasz KAISER, Manjunath KUDLUR, Josh LEVENBERG, Dandelion MANÉ, Rajat MONGA, Sherry MOORE, Derek MURRAY, Chris OLAH, Mike SCHUSTER, Jonathon SHLENS, Benoit STEINER, Ilya SUTSKEVER, Kunal TALWAR, Paul TUCKER, Vincent VANHOUCKE, Vijay VASUDEVAN, Fernanda VIÉGAS, Oriol VINYALS, Pete WARDEN, Martin WATTENBERG, Martin WICKE, Yuan YU et Xiaoqiang ZHENG : TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL <https://www.tensorflow.org/>. Software available from tensorflow.org.
- Alexandre ADAM, Adam COOGAN, Nikolay MALKIN, Ronan LEGIN, Laurence PERREAULT-LEVASSEUR, Yashar HEZAVEH et Yoshua BENGIO : Posterior samples of source galaxies in strong gravitational lenses with score-based priors. *arXiv e-prints*, page arXiv:2211.03812, novembre 2022.
- Alexandre ADAM, Laurence PERREAULT-LEVASSEUR, Yashar HEZAVEH et Max WELLING : Pixelated Reconstruction of Foreground Density and Background Surface Brightness in Gravitational Lensing Systems using Recurrent Inference Machines. *arXiv e-prints*, page arXiv:2301.04168, janvier 2023.
- Adam AMARA, R. Benton METCALF, Thomas J. COX et Jeremiah P. OSTRIKER : Simulations of strong gravitational lensing with substructure. *MNRAS*, 367(4):1367–1378, avril 2006.
- Noemi ANAU MONTEL, Adam COOGAN, Camila CORREA, Konstantin KARCHEV et Christoph WENIGER : Estimating the warm dark matter mass from strong lensing images with truncated marginal neural ratio estimation. *MNRAS*, 518(2):2746–2760, janvier 2023.
- Garry W. ANGUS, Huan Yuan SHAN, Hong Sheng ZHAO et Benoit FAMAÉY : On the proof of dark matter, the law of gravity, and the mass of neutrinos. *The Astrophysical Journal*, 654(1):L13, dec 2006. URL <https://dx.doi.org/10.1086/510738>.

ASTROPY COLLABORATION, A. M. PRICE-WHELAN, B. M. SIPÓCZ, H. M. GÜNTHER, P. L. LIM, S. M. CRAWFORD, S. CONSEIL, D. L. SHUPE, M. W. CRAIG, N. DENCHEVA, A. GINSBURG, J. T. VANDERPLAS, L. D. BRADLEY, D. PÉREZ-SUÁREZ, M. DE VAL-BORRO, T. L. ALDCROFT, K. L. CRUZ, T. P. ROBITAILLE, E. J. TOLLERUD, C. ARDELEAN, T. BABEJ, Y. P. BACH, M. BACHETTI, A. V. BAKANOV, S. P. BAMFORD, G. BARENTSEN, P. BARMBY, A. BAUMBACH, K. L. BERRY, F. BISCANI, M. BOQUIEN, K. A. BOSTROEM, L. G. BOUMA, G. B. BRAMMER, E. M. BRAY, H. BREYTENBACH, H. BUDELMEIJER, D. J. BURKE, G. CALDERONE, J. L. CANO RODRÍGUEZ, M. CARA, J. V. M. CARDOSO, S. CHEEDELLA, Y. COPIN, L. CORRALES, D. CRICHTON, D. D'AVELLA, C. DEIL, É. DEPAGNE, J. P. DIETRICH, A. DONATH, M. DROETTBOOM, N. EARL, T. ERBEN, S. FABBRO, L. A. FERREIRA, T. FINETHY, R. T. FOX, L. H. GARRISON, S. L. J. GIBBONS, D. A. GOLDSTEIN, R. GOMMERS, J. P. GRECO, P. GREENFIELD, A. M. GROENER, F. GROLLIER, A. HAGEN, P. HIRST, D. HOMEIER, A. J. HORTON, G. HOSSEINZADEH, L. HU, J. S. HUNKELER, Ž. IVEZIĆ, A. JAIN, T. JENNESS, G. KANAREK, S. KENDREW, N. S. KERN, W. E. KERZENDORF, A. KHVALKO, J. KING, D. KIRKBY, A. M. KULKARNI, A. KUMAR, A. LEE, D. LENZ, S. P. LITTLEFAIR, Z. MA, D. M. MACLEOD, M. MASTROPIETRO, C. MCCULLY, S. MONTAGNAC, B. M. MORRIS, M. MUELLER, S. J. MUMFORD, D. MUNA, N. A. MURPHY, S. NELSON, G. H. NGUYEN, J. P. NINAN, M. NÖTHER, S. OGAZ, S. OH, J. K. PARREJKO, N. PARLEY, S. PASCUAL, R. PATIL, A. A. PATIL, A. L. PLUNKETT, J. X. PROCHASKA, T. RASTOGI, V. REDDY JANGA, J. SABATER, P. SAKURIKAR, M. SEIFERT, L. E. SHERBERT, H. SHERWOOD-TAYLOR, A. Y. SHIH, J. SICK, M. T. SILBINGER, S. SINGANAMALLA, L. P. SINGER, P. H. SLADEN, K. A. SOOLEY, S. SORNARAJAH, O. STREICHER, P. TEUBEN, S. W. THOMAS, G. R. TREMBLAY, J. E. H. TURNER, V. TERRÓN, M. H. VAN KERKWIJK, A. DE LA VEGA, L. L. WATKINS, B. A. WEAVER, J. B. WHITMORE, J. WOILLET, V. ZABALZA et ASTROPY CONTRIBUTORS : The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. *AJ*, 156(3):123, septembre 2018.

ASTROPY COLLABORATION, Adrian M. PRICE-WHELAN, Pey Lian LIM, Nicholas EARL, Nathaniel STARKMAN, Larry BRADLEY, David L. SHUPE, Aarya A. PATIL, Lia CORRALES, C. E. BRASSEUR, Maximilian NÖTHER, Axel DONATH, Erik TOLLERUD, Brett M. MORRIS, Adam GINSBURG, Eero VAHER, Benjamin A. WEAVER, James TOCKNELL, William JAMIESON, Marten H. VAN KERKWIJK, Thomas P. ROBITAILLE, Bruce MERRY, Matteo BACHETTI, H. Moritz GÜNTHER, Thomas L. ALDCROFT, Jaime A. ALVARADO-MONTES, Anne M. ARCHIBALD, Attila BÓDI, Shreyas BAPAT, Geert BARENTSEN, Juanjo BAZÁN, Manish BISWAS, Médéric BOQUIEN, D. J. BURKE, Daria CARA, Mihai CARA, Kyle E. CONROY, Simon CONSEIL, Matthew W. CRAIG, Robert M. CROSS, Kelle L. CRUZ, Francesco D'EUGENIO, Nadia DENCHEVA, Hadrien A. R. DEVILLEPOIX,

Jörg P. DIETRICH, Arthur Davis EIGENBROT, Thomas ERBEN, Leonardo FERREIRA, Daniel FOREMAN-MACKEY, Ryan FOX, Nabil FREIJ, Suyog GARG, Robel GEDA, Lauren GLATTLY, Yash GONDHALKAR, Karl D. GORDON, David GRANT, Perry GREENFIELD, Austen M. GROENER, Steve GUEST, Sebastian GUROVICH, Rasmus HANDBERG, Akeem HART, Zac HATFIELD-DODDS, Derek HOMEIER, Griffin HOSSEINZADEH, Tim JENNESS, Craig K. JONES, Prajwel JOSEPH, J. Bryce KALMBACH, Emir KARAMEHMETOGLU, Mikołaj KAŁUSZYŃSKI, Michael S. P. KELLEY, Nicholas KERN, Wolfgang E. KERZENDORF, Eric W. KOCH, Shankar KULUMANI, Antony LEE, Chun LY, Zhiyuan MA, Conor MACBRIDE, Jakob M. MALJAARS, Demitri MUNA, N. A. MURPHY, Henrik NORMAN, Richard O’STEEN, Kyle A. OMAN, Camilla PACIFICI, Sergio PASCUAL, J. PASCUAL-GRANADO, Rohit R. PATIL, Gabriel I. PERREN, Timothy E. PICKERING, Tanuj RASTOGI, Benjamin R. ROULSTON, Daniel F. RYAN, Eli S. RYKOFF, Jose SABATER, Parikshit SAKURIKAR, Jesús SALGADO, Aniket SANGHI, Nicholas SAUNDERS, Volodymyr SAVCHENKO, Ludwig SCHWARDT, Michael SEIFERT-ECKERT, Albert Y. SHIH, Anany Shrey JAIN, Gyanendra SHUKLA, Jonathan SICK, Chris SIMPSON, Sudheesh SINGANAMALLA, Leo P. SINGER, Jaladh SINGHAL, Manodeep SINHA, Brigitta M. SIPIŃCZ, Lee R. SPITLER, David STANSBY, Ole STREICHER, Jani ŠUMAK, John D. SWINBANK, Dan S. TARANU, Nikita TEWARY, Grant R. TREMBLAY, Miguel DE VAL-BORRO, Samuel J. VAN KOOTEN, Zlatan VASOVIĆ, Shresth VERMA, José Vinícius DE MIRANDA CARDOSO, Peter K. G. WILLIAMS, Tom J. WILSON, Benjamin WINKEL, W. M. WOOD-VASEY, Rui XUE, Peter YOACHIM, Chen ZHANG, Andrea ZONCA et ASTROPY PROJECT CONTRIBUTORS : The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package. *ApJ*, 935(2):167, août 2022.

ASTROPY COLLABORATION, Thomas P. ROBITAILLE, Erik J. TOLLERUD, Perry GREENFIELD, Michael DROETTBOOM, Erik BRAY, Tom ALDCROFT, Matt DAVIS, Adam GINSBURG, Adrian M. PRICE-WHELAN, Wolfgang E. KERZENDORF, Alexander CONLEY, Neil CRIGHTON, Kyle BARBARY, Demitri MUNA, Henry FERGUSON, Frédéric GROLLIER, Madhura M. PARIKH, Prasanth H. NAIR, Hans M. UNTHER, Christoph DEIL, Julien WOILLET, Simon CONSEIL, Roban KRAMER, James E. H. TURNER, Leo SINGER, Ryan FOX, Benjamin A. WEAVER, Victor ZABALZA, Zachary I. EDWARDS, K. AZALEE BOSTROEM, D. J. BURKE, Andrew R. CASEY, Steven M. CRAWFORD, Nadia DENCHEVA, Justin ELY, Tim JENNESS, Kathleen LABRIE, Pey Lian LIM, Francesco PIERFEDERICI, Andrew PONTZEN, Andy PTAK, Brian REFSDAL, Mathieu SERVILLAT et Ole STREICHER : Astropy: A community Python package for astronomy. *A&A*, 558:A33, octobre 2013.

- Jimmy Lei BA, Jamie Ryan KIROS et Geoffrey E. HINTON : Layer Normalization. *arXiv e-prints*, page arXiv:1607.06450, juillet 2016.
- Edward A. BALTZ, Phil MARSHALL et Masamune OGURI : Analytic models of plausible gravitational lens potentials. *J. Cosmology Astropart. Phys.*, 2009(1):015, janvier 2009.
- Indranil BANIK, Charalambos PITTORDIS, Will SUTHERLAND, Benoit FAMAHEY, Rodrigo IBATA, Steffen MIESKE et Hongsheng ZHAO : Strong constraints on the gravitational law from Gaia DR3 wide binaries. *Monthly Notices of the Royal Astronomical Society*, 527(3):4573–4615, 11 2023. ISSN 0035-8711. URL <https://doi.org/10.1093/mnras/stad3393>.
- Rennan BAR-KANA : Effect of Large-Scale Structure on Multiply Imaged Sources. *ApJ*, 468:17, septembre 1996.
- E. E. BARNARD et A. C. RANYARD : Structure of the Milky Way. *Knowledge: An Illustrated Magazine of Science*, 17:253, novembre 1894.
- Peter S. BEHROOZI, Risa H. WECHSLER et Charlie CONROY : The Average Star Formation Histories of Galaxies in Dark Matter Halos from  $z = 0-8$ . *ApJ*, 770(1):57, juin 2013.
- F. W. BESSEL : On the variations of the proper motions of Procyon and Sirius. *MNRAS*, 6:136–141, décembre 1844.
- Simon BIRRER et Adam AMARA : lenstronomy: Multi-purpose gravitational lens modelling software package. *Physics of the Dark Universe*, 22:189–201, 2018. ISSN 2212-6864. URL <https://www.sciencedirect.com/science/article/pii/S2212686418301869>.
- Simon BIRRER, Adam AMARA et Alexandre REFREGIER : Lensing substructure quantification in RXJ1131-1231: a 2 keV lower bound on dark matter thermal relic mass. *J. Cosmology Astropart. Phys.*, 2017(5):037, mai 2017a.
- Simon BIRRER, Anowar J. SHAJIB, Daniel GILMAN, Aymeric GALAN, Jelle AALBERS, Martin MILLON, Robert MORGAN, Giulia PAGANO, Ji Won PARK, Luca TEODORI, Nicolas TESSORE, Madison UELAND, Lyne Van de VYVERE, Sebastian WAGNER-CARENA, Ewoud WEMPE, Lilan YANG, Xuheng DING, Thomas SCHMIDT, Dominique SLUSE, Ming ZHANG et Adam AMARA : lenstronomy ii: A gravitational lensing software ecosystem. *Journal of Open Source Software*, 6 (62):3283, 2021. URL <https://doi.org/10.21105/joss.03283>.
- Simon BIRRER, Cyril WELSCHEN, Adam AMARA et Alexandre REFREGIER : Line-of-sight effects in strong lensing: putting theory into practice. *J. Cosmology Astropart. Phys.*, 2017(4):049, avril 2017b.
- Christopher M BISHOP et Nasser M NASRABADI : *Pattern recognition and machine learning*, volume 4. Springer, 2006.

- Roger BLANDFORD et Ramesh NARAYAN : Fermat's Principle, Caustics, and the Classification of Gravitational Lens Images. *ApJ*, 310:568, novembre 1986.
- Johann BREHMER, Siddharth MISHRA-SHARMA, Joeri HERMANS, Gilles LOUPPE et Kyle CRANMER : Mining for Dark Matter Substructure: Inferring Subhalo Population Properties from Strong Lenses with Machine Learning. *ApJ*, 886(1):49, novembre 2019.
- James S. BULLOCK et Michael BOYLAN-KOLCHIN : Small-scale challenges to the  $\Lambda$ CDM paradigm. *Annual Review of Astronomy and Astrophysics*, 55(1):343–387, 2017. URL <https://doi.org/10.1146/annurev-astro-091916-055313>.
- Sean M. CARROLL, William H. PRESS et Edwin L. TURNER : The cosmological constant. *ARA&A*, 30:499–542, janvier 1992.
- Jacqueline CHEN, Andrey V. KRAVTSOV et Charles R. KEETON : Lensing Optical Depths for Substructure and Isolated Dark Matter Halos. *ApJ*, 592(1):24–31, juillet 2003.
- Masashi CHIBA : Probing Dark Matter Substructure in Lens Galaxies. *ApJ*, 565(1):17–23, janvier 2002.
- Douglas CLOWE, Maruša BRADAČ, Anthony H. GONZALEZ, Maxim MARKEVITCH, Scott W. RANDALL, Christine JONES et Dennis ZARITSKY : A Direct Empirical Proof of the Existence of Dark Matter. *ApJ*, 648(2):L109–L113, septembre 2006.
- Douglas CLOWE, Anthony GONZALEZ et Maxim MARKEVITCH : Weak-Lensing Mass Reconstruction of the Interacting Cluster 1E 0657-558: Direct Evidence for the Existence of Dark Matter. *ApJ*, 604(2):596–603, avril 2004.
- Thomas E. COLLETT : The Population of Galaxy-Galaxy Strong Lenses in Forthcoming Optical Imaging Surveys. *ApJ*, 811(1):20, septembre 2015.
- Andrew COLLETTE, Thomas KLUYVER, Thomas A CASWELL, James TOCKNELL, Jerome KIEFFER, Aleksandar JELENAK, Anthony SCOPATZ, Darren DALE, CHEN, Thomas VINCENT, Matt EINHORN, PAYNO, JULIAGARRIGA, Pierlauro SCIARELLI, Valentin VALLS, Satrajit GHOSH, Ulrik Kofoed PEDERSEN, Mark KITTISOPIKUL, JAKIRKHAM, Martin RASPAUD, Cyril DANILEVSKI, Hameer ABBASI, John READEY, Kai MÜHLBAUER, Andrey PARAMONOV, Lawrence CHAN, Robin De SCHEPPER, V. Armando SOLÉ, JIALIN et Daniel Hay GUEST : h5py/h5py: 3.8.0-aarch64-wheels, janvier 2023. URL <https://doi.org/10.5281/zenodo.7568214>.
- Arthur B. CONGDON et Charles R. KEETON : *Principles of Gravitational Lensing: Light Deflection as a Probe of Astrophysics and Cosmology*. 2018.
- Adam COOGAN, Konstantin KARCHEV et Christoph WENIGER : Targeted Likelihood-Free Inference of Dark Matter Substructure in Strongly-Lensed Galaxies. *arXiv e-prints*, page arXiv:2010.07032, octobre 2020.

- Kyle CRANMER, Johann BREHMER et Gilles LOUPPE : The frontier of simulation-based inference. *Proceedings of the National Academy of Science*, 117(48):30055–30062, décembre 2020.
- Atınc Çağan ŞENGÜL, Cora DVORKIN, Bryan OSTDIEK et Arthur TSANG : Substructure detection reanalysed: dark perturber shown to be a line-of-sight halo. *MNRAS*, 515(3):4391–4401, septembre 2022.
- Atınc Çağan ŞENGÜL, Arthur TSANG, Ana DIAZ RIVERO, Cora DVORKIN, Hong-Ming ZHU et Uroš SELJAK : Quantifying the line-of-sight halo contribution to the dark matter convergence power spectrum from strong gravitational lenses. *Phys. Rev. D*, 102(6):063502, septembre 2020.
- George CYBENKO : Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.
- Francis-Yan CYR-RACINE, Leonidas A. MOUSTAKAS, Charles R. KEETON, Kris SIGURDSON et Daniel A. GILMAN : Dark census: Statistically detecting the satellite populations of distant galaxies. *Phys. Rev. D*, 94(4):043505, août 2016.
- Giulia DESPALI et Simona VEGETTI : The impact of baryonic physics on the subhalo mass function and implications for gravitational lensing. *MNRAS*, 469(2):1997–2010, août 2017.
- Giulia DESPALI, Simona VEGETTI, Simon D. M. WHITE, Carlo GIOCOLI et Frank C. VAN DEN BOSCH : Modelling the line-of-sight contribution in substructure lensing. *MNRAS*, 475(4):5424–5442, avril 2018.
- Birendra DHANASINGHAM, Francis-Yan CYR-RACINE, Annika H G PETER, Andrew BENSON et Daniel GILMAN : Interlopers speak out: studying the dark universe using small-scale lensing anisotropies. *Monthly Notices of the Royal Astronomical Society*, 518(4):5843–5861, oct 2022. URL <https://doi.org/10.1093/mnras/stac2993>.
- Benedikt DIEMER : COLOSSUS: A Python Toolkit for Cosmology, Large-scale Structure, and Dark Matter Halos. *ApJS*, 239(2):35, décembre 2018.
- A. DRLICA-WAGNER, K. BECHTOL, E. S. RYKOFF, E. LUQUE, A. QUEIROZ, Y. Y. MAO, R. H. WECHSLER, J. D. SIMON, B. SANTIAGO, B. YANNY, E. BALBINOT, S. DODELSON, A. FAUSTI NETO, D. J. JAMES, T. S. LI, M. A. G. MAIA, J. L. MARSHALL, A. PIERES, K. STRINGER, A. R. WALKER, T. M. C. ABBOTT, F. B. ABDALLA, S. ALLAM, A. BENOIT-LÉVY, G. M. BERNSTEIN, E. BERTIN, D. BROOKS, E. BUCKLEY-GEER, D. L. BURKE, A. CARNERO ROSELL, M. CARRASCO KIND, J. CARRETERO, M. CROCCE, L. N. DA COSTA, S. DESAI, H. T. DIEHL, J. P. DIETRICH, P. DOEL, T. F. EIFLER, A. E. EVRARD, D. A. FINLEY, B. FLAUGHER, P. FOSALBA, J. FRIEMAN, E. GAZTANAGA, D. W. GERDES, D. GRUEN, R. A. GRUENDL, G. GUTIERREZ, K. HONSCHEID, K. KUEHN, N. KUROPATKIN, O. LAHAV, P. MARTINI, R. MIQUEL, B. NORD, R. OGANDO, A. A. PLAZAS, K. REIL, A. ROODMAN, M. SAKO, E. SANCHEZ, V. SCARPINE, M. SCHUBNELL,

- I. SEVILLA-NOARBE, R. C. SMITH, M. SOARES-SANTOS, F. SOBREIRA, E. SUCHYTA, M. E. C. SWANSON, G. TARLE, D. TUCKER, V. VIKRAM, W. WESTER, Y. ZHANG, J. ZUNTZ et DES COLLABORATION : Eight Ultra-faint Galaxy Candidates Discovered in Year Two of the Dark Energy Survey. *ApJ*, 813(2):109, novembre 2015.
- Alan R. DUFFY, Joop SCHAYE, Scott T. KAY, Claudio DALLA VECCHIA, Richard A. BATTYE et C. M. BOOTH : Impact of baryon physics on dark matter structures: a detailed simulation study of halo density profiles. *MNRAS*, 405(4):2161–2178, juillet 2010.
- Jaan EINASTO, Ants KAASIK et Enn SAAR : Dynamic evidence on massive coronas of galaxies. *Nature*, 250(5464):309–310, juillet 1974.
- Albert EINSTEIN : Lens-Like Action of a Star by the Deviation of Light in the Gravitational Field. *Science*, 84(2188):506–507, décembre 1936.
- Ross FADELY et Charles R. KEETON : Substructure in the lens HE 0435-1223. *MNRAS*, 419(2):936–951, janvier 2012.
- Pierre FLEURY, Julien LARENA et Jean-Philippe UZAN : Gravitational lenses in arbitrary spacetimes. *Classical and Quantum Gravity*, 38(8):085002, avril 2021a.
- Pierre FLEURY, Julien LARENA et Jean-Philippe UZAN : Line-of-sight effects in strong gravitational lensing. *J. Cosmology Astropart. Phys.*, 2021(8):024, août 2021b.
- GAIA COLLABORATION, A. VALLENARI, A. G. A. BROWN, T. PRUSTI, J. H. J. DE BRUIJNE, F. ARENOU, C. BABUSIAUX, M. BIERMANN, O. L. CREEVEY, C. DUCOURANT, D. W. EVANS, L. EYER, R. GUERRA, A. HUTTON, C. JORDI, S. A. KLIONER, U. L. LAMMERS, L. LINDEGREN, X. LURI, F. MIGNARD, C. PANEM, D. POURBAIX, S. RANDICH, P. SARTORETTI, C. SOUBIRAN, P. TANGA, N. A. WALTON, C. A. L. BAILER-JONES, U. BASTIAN, R. DRIMMEL, F. JANSEN, D. KATZ, M. G. LATTANZI, F. VAN LEEUWEN, J. BAKKER, C. CACCIARI, J. CASTAÑEDA, F. DE ANGELI, C. FABRICIUS, M. FOUESNEAU, Y. FRÉMAT, L. GALLUCCIO, A. GUERRIER, U. HELLTER, E. MASANA, R. MESSINEO, N. MOWLAVI, C. NICOLAS, K. NIENARTOWICZ, F. PAILLER, P. PANUZZO, F. RICLET, W. ROUX, G. M. SEABROKE, R. SORDO, F. THÉVENIN, G. GRACIA-ABRIL, J. PORTELL, D. TEYSSIER, M. ALTMANN, R. ANDRAE, M. AUDARD, I. BELLAS-VELIDIS, K. BENSON, J. BERTHIER, R. BLOMME, P. W. BURGESS, D. BUSONERO, G. BUSSO, H. CÁNOVAS, B. CARRY, A. CELLINO, N. CHEEK, G. CLEMENTINI, Y. DAMERDJI, M. DAVIDSON, P. DE TEODORO, M. NUÑEZ CAMPOS, L. DELCHAMBRE, A. DELL’ORO, P. ESQUEJ, J. FERNÁNDEZ-HERNÁNDEZ, E. FRAILE, D. GARABATO, P. GARCÍA-LARIO, E. GOSSET, R. HAIGRON, J. L. HALBWACHS, N. C. HAMBLY, D. L. HARRISON, J. HERNÁNDEZ, D. HESTROFFER, S. T. HODGKIN, B. HOLL, K. JANSSEN, G. JEVARDAT DE FOMBELLE, S. JORDAN, A. KRONE-MARTINS,

A. C. LANZAFAME, W. LÖFFLER, O. MARCHAL, P. M. MARRESE, A. MOITINHO, K. MUI-  
NONEN, P. OSBORNE, E. PANCINO, T. PAUWELS, A. RECIO-BLANCO, C. REYLÉ, M. RIELLO,  
L. RIMOLDINI, T. ROEGIERS, J. RYBIZKI, L. M. SARRO, C. SIOPIS, M. SMITH, A. SOZZETTI,  
E. UTRILLA, M. VAN LEEUWEN, U. ABBAS, P. ÁBRAHÁM, A. ABREU ARAMBURU, C. AERTS,  
J. J. AGUADO, M. AJAJ, F. ALDEA-MONTERO, G. ALTAVILLA, M. A. ÁLVAREZ, J. ALVES,  
F. ANDERS, R. I. ANDERSON, E. ANGLADA VARELA, T. ANTOJA, D. BAINES, S. G. BAKER,  
L. BALAGUER-NÚÑEZ, E. BALBINOT, Z. BALOG, C. BARACHE, D. BARBATO, M. BARROS,  
M. A. BARSTOW, S. BARTOLOMÉ, J. L. BASSILANA, N. BAUCHET, U. BECCIANI, M. BEL-  
LAZZINI, A. BERIHUETE, M. BERNET, S. BERTONE, L. BIANCHI, A. BINNENFELD, S. BLANCO-  
CUARESMA, A. BLAZERE, T. BOCH, A. BOMBRUN, D. BOSSINI, S. BOUQUILLON, A. BRAGAGLIA,  
L. BRAMANTE, E. BREEDT, A. BRESSAN, N. BROUILLET, E. BRUGALETTA, B. BUCCIARELLI,  
A. BURLACU, A. G. BUTKEVICH, R. BUZZI, E. CAFFAU, R. CANCELLIERE, T. CANTAT-GAUDIN,  
R. CARBALLO, T. CARLUCCI, M. I. CARNERERO, J. M. CARRASCO, L. CASAMIQUELA, M. CAS-  
TELLANI, A. CASTRO-GINARD, L. CHAOWL, P. CHARLOT, L. CHEMIN, V. CHIARAMIDA, A. CHIA-  
VASSA, N. CHORNAY, G. COMORETTO, G. CONTURSI, W. J. COOPER, T. CORNEZ, S. COWELL,  
F. CRIFO, M. CROPPER, M. CROSTA, C. CROWLEY, C. DAFONTE, A. DAPERGOLAS, M. DA-  
VID, P. DAVID, P. DE LAVERNY, F. DE LUISE, R. DE MARCH, J. DE RIDDER, R. DE SOUZA,  
A. DE TORRES, E. F. DEL PELOSO, E. DEL POZO, M. DELBO, A. DELGADO, J. B. DELISLE,  
C. DEMOUCHY, T. E. DHARMAWARDENA, P. DI MATTEO, S. DIAKITE, C. DIENER, E. DISTE-  
FANO, C. DOLDING, B. EDVARDSSON, H. ENKE, C. FABRE, M. FABRIZIO, S. FAIGLER, G. FE-  
DORETS, P. FERNIQUE, A. FIENGA, F. FIGUERAS, Y. FOURNIER, C. FOURON, F. FRAGKOURI,  
M. GAI, A. GARCIA-GUTIERREZ, M. GARCIA-REINALDOS, M. GARCÍA-TORRES, A. GARO-  
FALO, A. GAVEL, P. GAVRAS, E. GERLACH, R. GEYER, P. GIACOBBE, G. GILMORE, S. GIRONA,  
G. GIUFFRIDA, R. GOMEL, A. GOMEZ, J. GONZÁLEZ-NÚÑEZ, I. GONZÁLEZ-SANTAMARÍA, J. J.  
GONZÁLEZ-VIDAL, M. GRANVIK, P. GUILLOUT, J. GUIRAUD, R. GUTIÉRREZ-SÁNCHEZ, L. P.  
GUY, D. HATZIDIMITRIOU, M. HAUSER, M. HAYWOOD, A. HELMER, A. HELMI, M. H. SAR-  
MIENTO, S. L. HIDALGO, T. HILGER, N. HŁADCZUK, D. HOBBS, G. HOLLAND, H. E. HUCKLE,  
K. JARDINE, G. JASNIEWICZ, A. JEAN-ANTOINE PICCOLO, Ó. JIMÉNEZ-ARRANZ, A. JORISSEN,  
J. JUARISTI CAMPILLO, F. JULBE, L. KARBEVSKA, P. KERVELLA, S. KHANNA, M. KONTIZAS,  
G. KORDOPATIS, A. J. KORN, Á. KÓSPÁL, Z. KOSTRZEWA-RUTKOWSKA, K. KRUSZYŃSKA,  
M. KUN, P. LAIZEAU, S. LAMBERT, A. F. LANZA, Y. LASNE, J. F. LE CAMPION, Y. LE-  
BRETON, T. LEBZELTER, S. LECCIA, N. LECLERC, I. LECOEUR-TAIBI, S. LIAO, E. L. LICATA,  
H. E. P. LINDSTRØM, T. A. LISTER, E. LIVANOU, A. LOBEL, A. LORCA, C. LOUP, P. MADRERO  
PARDO, A. MAGDALENO ROMEO, S. MANAGAU, R. G. MANN, M. MANTEIGA, J. M. MARCHANT,



M. MARCONI, J. MARCOS, M. M. S. MARCOS SANTOS, D. MARÍN PINA, S. MARINONI, F. MAROCCO, D. J. MARSHALL, L. MARTIN POLO, J. M. MARTÍN-FLEITAS, G. MARTON, N. MARY, A. MASIP, D. MASSARI, A. MASTROBUONO-BATTISTI, T. MAZEH, P. J. McMILLAN, S. MESSINA, D. MICHALIK, N. R. MILLAR, A. MINTS, D. MOLINA, R. MOLINARO, L. MOLNÁR, G. MONARI, M. MONGUIÓ, P. MONTEGRIFFO, A. MONTERO, R. MOR, A. MORA, R. MORBIDELLI, T. MOREL, D. MORRIS, T. MURAVEVA, C. P. MURPHY, I. MUSELLA, Z. NAGY, L. NOVAL, F. OCAÑA, A. OGDEN, C. ORDENOVIC, J. O. OSINDE, C. PAGANI, I. PAGANO, L. PALAVERSA, P. A. PALICIO, L. PALLAS-QUINTELA, A. PANAHÍ, S. PAYNE-WARDENAAR, X. PEÑALOSA ESTELLER, A. PENTTILÄ, B. PICHON, A. M. PIERSIMONI, F. X. PINEAU, E. PLACHY, G. PLUM, E. POGGIO, A. PRŠA, L. PULONE, E. RACERO, S. RAGAINI, M. RAINER, C. M. RAITERI, N. RAMBAUX, P. RAMOS, M. RAMOS-LERATE, P. RE FIORENTIN, S. REGIBO, P. J. RICHARDS, C. RIOS DIAZ, V. RIPEPI, A. RIVA, H. W. RIX, G. RIXON, N. ROBICHON, A. C. ROBIN, C. ROBIN, M. ROELEN, H. R. O. ROGUES, L. ROHRBASSER, M. ROMERO-GÓMEZ, N. ROWELL, F. ROYER, D. RUZ MIERES, K. A. RYBICKI, G. SADOWSKI, A. SÁEZ NÚÑEZ, A. SAGRISTÀ SELLÉS, J. SAHLMANN, E. SALGUERO, N. SAMARAS, V. SANCHEZ GIMENEZ, N. SANNA, R. SANTOVEÑA, M. SARASSO, M. SCHULTHEIS, E. SCIACCA, M. SEGOL, J. C. SEGOVIA, D. SÉGRANSAN, D. SEMEUX, S. SHAHAF, H. I. SIDDIQUI, A. SIEBERT, L. SILTALA, A. SILVELO, E. SLEZAK, I. SLEZAK, R. L. SMART, O. N. SNAITH, E. SOLANO, F. SOLITRO, D. SOUAMI, J. SOUCHAY, A. SPAGNA, L. SPINA, F. SPOTO, I. A. STEELE, H. STEIDELMÜLLER, C. A. STEPHENSON, M. SÜVEGES, J. SURDEJ, L. SZABADOS, E. SZEGEDI-ELEK, F. TARIS, M. B. TAYLOR, R. TEIXEIRA, L. TOLOMEI, N. TONELLO, F. TORRA, J. TORRA, G. TORRALBA ELIPE, M. TRABUCCHI, A. T. TSOUNIS, C. TURON, A. ULLA, N. UNGER, M. V. VAILLANT, E. VAN DILLEN, W. VAN REEVEN, O. VANEL, A. VECCHIATO, Y. VIALA, D. VICENTE, S. VOUTSINAS, M. WEILER, T. WEVERS, Ł. WYRZYKOWSKI, A. YOLDAS, P. YVARD, H. ZHAO, J. ZOREC, S. ZUCKER et T. ZWITTER : Gaia Data Release 3. Summary of the content and survey properties. *A&A*, 674:A1, juin 2023.

Jonathan P. GARDNER, John C. MATHER, Mark CLAMPIN, Rene DOYON, Matthew A. GREENHOUSE, Heidi B. HAMMEL, John B. HUTCHINGS, Peter JAKOBSEN, Simon J. LILLY, Knox S. LONG, Jonathan I. LUNINE, Mark J. McCAUGHREAN, Matt MOUNTAIN, John NELLA, George H. RIEKE, Marcia J. RIEKE, Hans-Walter RIX, Eric P. SMITH, George SONNEBORN, Massimo STIAVELLI, H. S. STOCKMAN, Rogier A. WINDHORST et Gillian S. WRIGHT : The James Webb Space Telescope. *Space Sci. Rev.*, 123(4):485–606, avril 2006.

Sebastiano GHIGNA, Ben MOORE, Fabio GOVERNATO, George LAKE, Thomas QUINN et Joachim STADEL : Dark matter haloes within clusters. *MNRAS*, 300(1):146–162, octobre 1998.

- Daniel GILMAN, Andrew BENSON, Jo BOVY, Simon BIRRER, Tommaso TREU et Anna NIERENBERG : The primordial matter power spectrum on sub-galactic scales. *MNRAS*, 512(3):3163–3188, mai 2022.
- Daniel GILMAN, Simon BIRRER, Anna NIERENBERG, Tommaso TREU, Xiaolong DU et Andrew BENSON : Warm dark matter chills out: constraints on the halo mass function and the free-streaming length of dark matter with eight quadruple-image strong gravitational lenses. *MNRAS*, 491(4):6077–6101, février 2020a.
- Daniel GILMAN, Simon BIRRER, Tommaso TREU, Charles R. KEETON et Anna NIERENBERG : Probing the nature of dark matter by forward modelling flux ratios in strong gravitational lenses. *MNRAS*, 481(1):819–834, novembre 2018.
- Daniel GILMAN, Simon BIRRER, Tommaso TREU, Anna NIERENBERG et Andrew BENSON : Probing dark matter structure down to  $10^7$  solar masses: flux ratio statistics in gravitational lenses with line-of-sight haloes. *MNRAS*, 487(4):5721–5738, août 2019.
- Daniel GILMAN, Jo BOVY, Tommaso TREU, Anna NIERENBERG, Simon BIRRER, Andrew BENSON et Omid SAMEIE : Strong lensing signatures of self-interacting dark matter in low-mass halos. *Monthly Notices of the Royal Astronomical Society*, 507:2432–2447, 5 2021. URL <http://arxiv.org/abs/2105.05259><http://dx.doi.org/10.1093/mnras/stab2335>.
- Daniel GILMAN, Xiaolong DU, Andrew BENSON, Simon BIRRER, Anna NIERENBERG et Tommaso TREU : Constraints on the mass-concentration relation of cold dark matter halos with 11 strong gravitational lenses. *MNRAS*, 492(1):L12–L16, février 2020b.
- Daniel Alejandro GILMAN : *Investigating the nature of dark matter with strong gravitational lensing*. Thèse de doctorat, University of California, Los Angeles, janvier 2020.
- Ian GOODFELLOW, Yoshua BENGIO et Aaron COURVILLE : *Deep learning*. MIT press, 2016.
- Charles R. HARRIS, K. Jarrod MILLMAN, Stéfan J. van der WALT, Ralf GOMMERS, Pauli VIRTANEN, David COURNAPEAU, Eric WIESER, Julian TAYLOR, Sebastian BERG, Nathaniel J. SMITH, Robert KERN, Matti PICUS, Stephan HOYER, Marten H. van KERKWIJK, Matthew BRETT, Allan HALDANE, Jaime Fernández del RÍO, Mark WIEBE, Pearu PETERSON, Pierre GÉRARD-MARCHANT, Kevin SHEPPARD, Tyler REDDY, Warren WECKESSER, Hameer ABBASI, Christoph GOHLKE et Travis E. OLIPHANT : Array programming with NumPy. *Nature*, 585(7825):357–362, septembre 2020. URL <https://doi.org/10.1038/s41586-020-2649-2>.
- Yashar D. HEZAVEH, Neal DALAL, Daniel P. MARRONE, Yao-Yuan MAO, Warren MORNINGSTAR, Di WEN, Roger D. BLANDFORD, John E. CARLSTROM, Christopher D. FASSNACHT, Gilbert P. HOLDER, Athol KEMBALL, Philip J. MARSHALL, Norman MURRAY, Laurence PERREAULT LEVASSEUR, Joaquin D. VIEIRA et Risa H. WECHSLER : Detection of Lensing Substructure Using

- ALMA Observations of the Dusty Galaxy SDP.81. *ApJ*, 823(1):37, mai 2016.
- Yashar D. HEZAVEH, Laurence PERREAULT LEVASSEUR et Philip J. MARSHALL : Fast automated analysis of strong gravitational lenses with convolutional neural networks. *Nature*, 548(7669):555–557, août 2017.
- Kurt HORNIK, Maxwell STINCHCOMBE et Halbert WHITE : Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989. ISSN 0893-6080. URL <https://www.sciencedirect.com/science/article/pii/0893608089900208>.
- J. W. HSUEH, W. ENZI, S. VEGETTI, M. W. AUGER, C. D. FASSNACHT, G. DESPALI, L. V. E. KOOPMANS et J. P. MCKEAN : SHARP - VII. New constraints on the dark matter free-streaming properties and substructure abundance from gravitationally lensed quasars. *MNRAS*, 492(2):3047–3059, février 2020.
- J. D. HUNTER : Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.
- Kaiki Taro INOUE, Ryuichi TAKAHASHI, Tomo TAKAHASHI et Tomoaki ISHIYAMA : Constraints on warm dark matter from weak lensing in anomalous quadruple lenses. *MNRAS*, 448(3):2704–2716, avril 2015.
- J. C. KAPTEYN : First Attempt at a Theory of the Arrangement and Motion of the Sidereal System. *ApJ*, 55:302, mai 1922.
- Konstantin KARCHEV, Adam COOGAN et Christoph WENIGER : Strong-lensing source reconstruction with variationally optimized Gaussian processes. *MNRAS*, 512(1):661–685, mai 2022.
- Diederik P KINGMA et Jimmy BA : Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- L. V. E. KOOPMANS : Gravitational imaging of cold dark matter substructures. *MNRAS*, 363(4):1136–1144, novembre 2005.
- L. V. E. KOOPMANS, I. W. A. BROWNE et N. J. JACKSON : Strong gravitational lensing with SKA. *New A Rev.*, 48(11-12):1085–1094, décembre 2004.
- Israel KOVNER : The Thick Gravitational Lens: A Lens Composed of Many Elements at Different Distances. *ApJ*, 316:52, mai 1987.
- Alexander LAROCHE, Daniel GILMAN, Xinyu LI, Jo BOVY et Xiaolong DU : Quantum fluctuations masquerade as haloes: bounds on ultra-light dark matter from quadruply imaged quasars. *MNRAS*, 517(2):1867–1883, décembre 2022.
- R. LAURELIS, J. AMIAUX, S. ARDUINI, J. L. AUGUÈRES, J. BRINCHMANN, R. COLE, M. CROPPER, C. DABIN, L. DUVET, A. EALET, B. GARILLI, P. GONDOIN, L. GUZZO, J. HOAR, H. HOEKSTRA, R. HOLMES, T. KITCHING, T. MACIASZEK, Y. MELLIER, F. PASIAN, W. PERCIVAL, J. RHODES,

G. SAAVEDRA CRIADO, M. SAUVAGE, R. SCARAMELLA, L. VALENZIANO, S. WARREN, R. BENDER, F. CASTANDER, A. CIMATTI, O. LE FÈVRE, H. KURKI-SUONIO, M. LEVI, P. LILJE, G. MEYLAN, R. NICHOL, K. PEDERSEN, V. POPA, R. REBOLO LOPEZ, H. W. RIX, H. ROTTGERING, W. ZEILINGER, F. GRUPP, P. HUDELLOT, R. MASSEY, M. MENEGHETTI, L. MILLER, S. PALTANI, S. PAULIN-HENRIKSSON, S. PIRES, C. SAXTON, T. SCHRABBACK, G. SEIDEL, J. WALSH, N. AGHANIM, L. AMENDOLA, J. BARTLETT, C. BACCIGALUPI, J. P. BEAULIEU, K. BENABED, J. G. CUBY, D. ELBAZ, P. FOSALBA, G. GAVAZZI, A. HELMI, I. HOOK, M. IRWIN, J. P. KNEIB, M. KUNZ, F. MANNUCCI, L. MOSCARDINI, C. TAO, R. TEYSSIER, J. WELLER, G. ZAMORANI, M. R. ZAPATERO OSORIO, O. BOULADE, J. J. FOUMOND, A. DI GIORGIO, P. GUTTRIDGE, A. JAMES, M. KEMP, J. MARTIGNAC, A. SPENCER, D. WALTON, T. BLÜMCHEN, C. BONOLI, F. BORTOLETTO, C. CERNA, L. CORCIONE, C. FABRON, K. JAHNKE, S. LIGORI, F. MADRID, L. MARTIN, G. MORGANTE, T. PAMPLONA, E. PRIETO, M. RIVA, R. TOLEDO, M. TRIFOGLIO, F. ZERBI, F. ABDALLA, M. DOUSPIS, C. GRENET, S. BORGANI, R. BOUWENS, F. COURBIN, J. M. DELOUIS, P. DUBATH, A. FONTANA, M. FRAILIS, A. GRAZIAN, J. KOPPENHÖFER, O. MANSUTTI, M. MELCHIOR, M. MIGNOLI, J. MOHR, C. NEISSNER, K. NODDLE, M. PONCET, M. SCODEGGIO, S. SERRANO, N. SHANE, J. L. STARCK, C. SURACE, A. TAYLOR, G. VERDOES-KLEIJN, C. VUERLI, O. R. WILLIAMS, A. ZACCHEI, B. ALTIERI, I. ESCUDERO SANZ, R. KOHLEY, T. OOSTERBROEK, P. ASTIER, D. BACON, S. BARDELLI, C. BAUGH, F. BELLAGAMBA, C. BENOIST, D. BIANCHI, A. BIVIANO, E. BRANCHINI, C. CARBONE, V. CARDONE, D. CLEMENTS, S. COLOMBI, C. CONSELICE, G. CRESCI, N. DEACON, J. DUNLOP, C. FEDELI, F. FONTANOT, P. FRANZETTI, C. GIOCOLI, J. GARCIA-BELLIDO, J. GOW, A. HEAVENS, P. HEWETT, C. HEYMANS, A. HOLLAND, Z. HUANG, O. ILBERT, B. JOACHIMI, E. JENNINS, E. KERINS, A. KIESSLING, D. KIRK, R. KOTAK, O. KRAUSE, O. LAHAV, F. VAN LEEUWEN, J. LESGOURGUES, M. LOMBARDI, M. MAGLIOCCHETTI, K. MAGUIRE, E. MAJEROTTO, R. MAOLI, F. MARULLI, S. MAUROGORDATO, H. MCCRACKEN, R. MCLURE, A. MELCHIORRI, A. MERSON, M. MORESCO, M. NONINO, P. NORBERG, J. PEACOCK, R. PELLO, M. PENNY, V. PETTORINO, C. DI PORTO, L. POZZETTI, C. QUERCELLINI, M. RADOVICH, A. RASSAT, N. ROCHE, S. RONAYETTE, E. ROSSETTI, B. SARTORIS, P. SCHNEIDER, E. SEMBOLONI, S. SERJEANT, F. SIMPSON, C. SKORDIS, G. SMADJA, S. SMARTT, P. SPANO, S. SPIRO, M. SULLIVAN, A. TILQUIN, R. TROTTA, L. VERDE, Y. WANG, G. WILLIGER, G. ZHAO, J. ZOUBIAN et E. ZUCCA : Euclid Definition Study Report. *arXiv e-prints*, page arXiv:1110.3193, octobre 2011.

Juho LEE, Yoonho LEE, Jungtaek KIM, Adam KOSIOREK, Seungjin CHOI et Yee Whye TEH : Set transformer: A framework for attention-based permutation-invariant neural networks. *In* Kamalika CHAUDHURI et Ruslan SALAKHUTDINOV, éditeurs : *Proceedings of the 36th International*

- Conference on Machine Learning*, volume 97 de *Proceedings of Machine Learning Research*, pages 3744–3753. PMLR, 09–15 Jun 2019. URL <https://proceedings.mlr.press/v97/lee19d.html>.
- Ronan LEGIN, Yashar HEZAVEH, Laurence PERREAULT LEVASSEUR et Benjamin WANDELT : Simulation-Based Inference of Strong Gravitational Lensing Parameters. *arXiv e-prints*, page arXiv:2112.05278, décembre 2021.
- Ronan LEGIN, Connor STONE, Yashar HEZAVEH et Laurence PERREAULT-LEVASSEUR : Population-Level Inference of Strong Gravitational Lenses with Neural Network-Based Selection Correction. *arXiv e-prints*, page arXiv:2207.04123, juillet 2022.
- Ran LI, Carlos S. FRENK, Shaun COLE, Qiao WANG et Liang GAO : Projection effects in the strong lensing study of subhaloes. *MNRAS*, 468(2):1426–1432, juin 2017.
- LORD KELVIN : *Baltimore Lectures on Molecular Dynamics and the Wave Theory of Light*. C.J. Clay & Sons, Cambridge University Press, 1904.
- Mark R. LOVELL, Carlos S. FRENK, Vincent R. EKE, Adrian JENKINS, Liang GAO et Tom THEUNS : The properties of warm dark matter haloes. *MNRAS*, 439(1):300–317, mars 2014.
- LSST DARK ENERGY SCIENCE COLLABORATION (LSST DESC), Bela ABOLFATHI, David ALONSO, Robert ARMSTRONG, Éric AUBOURG, Humna AWAN, Yadu N. BABUJI, Franz Erik BAUER, Rachel BEAN, George BECKETT, Rahul BISWAS, Joanne R. BOGART, Dominique BOUTIGNY, Kyle CHARD, James CHIANG, Chuck F. CLAVER, Johann COHEN-TANUGI, Céline COMBET, Andrew J. CONNOLLY, Scott F. DANIEL, Seth W. DIGEL, Alex DRLICA-WAGNER, Richard DUBOIS, Emmanuel GANGLER, Eric GAWISER, Thomas GLANZMAN, Phillipe GRIS, Salman HABIB, Andrew P. HEARIN, Katrin HEITMANN, Fabio HERNANDEZ, Renée HLOŽEK, Joseph HOLLOWED, Mustapha ISHAK, Željko IVEZIĆ, Mike JARVIS, Saurabh W. JHA, Steven M. KAHN, J. Bryce KALMBACH, Heather M. KELLY, Eve KOVACS, Danila KORYTOV, K. Simon KRUGHOFF, Craig S. LAGE, François LANUSSE, Patricia LARSEN, Laurent LE GUILLOU, Nan LI, Emily Phillips LONGLEY, Robert H. LUPTON, Rachel MANDELBAUM, Yao-Yuan MAO, Phil MARSHALL, Joshua E. MEYERS, Marc MONIEZ, Christopher B. MORRISON, Andrei NOMEROTSKI, Paul O’CONNOR, HyeYun PARK, Ji Won PARK, Julien PELOTON, Daniel PERREFORT, James PERRY, Stéphane PLASZCZYNSKI, Adrian POPE, Andrew RASMUSSEN, Kevin REIL, Aaron J. ROODMAN, Eli S. RYKOFF, F. Javier SÁNCHEZ, Samuel J. SCHMIDT, Daniel SCOLNIC, Christopher W. STUBBS, J. Anthony TYSON, Thomas D. URAM, Antonia Sierra VILLARREAL, Christopher W. WALTER, Matthew P. WIESNER, W. Michael WOOD-VASEY et Joe ZUNTZ : The LSST DESC DC2 Simulated Sky Survey. *ApJS*, 253(1):31, mars 2021.
- Chelsea L. MACLEOD, Ramsey JONES, Eric AGOL et Christopher S. KOCHANEK : Detection of Substructure in the Gravitationally Lensed Quasar MG0414+0534 Using Mid-infrared and Radio

- VLBI Observations. *ApJ*, 773(1):35, août 2013.
- M. MARKEVITCH, A. H. GONZALEZ, L. DAVID, A. VIKHLININ, S. MURRAY, W. FORMAN, C. JONES et W. TUCKER : A Textbook Example of a Bow Shock in the Merging Galaxy Cluster 1E 0657-56. *ApJ*, 567(1):L27–L31, mars 2002.
- Curtis MCCULLY, Charles R. KEETON, Kenneth C. WONG et Ann I. ZABLUDOFF : A new hybrid framework to efficiently model lines of sight to gravitational lenses. *MNRAS*, 443(4):3631–3642, octobre 2014.
- Curtis MCCULLY, Charles R. KEETON, Kenneth C. WONG et Ann I. ZABLUDOFF : Quantifying Environmental and Line-of-sight Effects in Models of Strong Gravitational Lens Systems. *ApJ*, 836(1):141, février 2017.
- John F. MEEKINS, Gilbert FRITZ, Talbot A. CHUBB et H. FRIEDMAN : Physical Sciences: X-rays from the Coma Cluster of Galaxies. *Nature*, 231(5298):107–108, mai 1971.
- Massimo MENEGHETTI : *Introduction to Gravitational Lensing; With Python Examples*, volume 956. 2021.
- R. Benton METCALF : The Importance of Intergalactic Structure to Gravitationally Lensed Quasars. *ApJ*, 629(2):673–679, août 2005.
- R. Benton METCALF et HongSheng ZHAO : Flux Ratios as a Probe of Dark Substructures in Quadruple-Image Gravitational Lenses. *ApJ*, 567(1):L5–L8, mars 2002.
- Siddharth MISHRA-SHARMA et Ge YANG : Strong Lensing Source Reconstruction Using Continuous Neural Fields. *arXiv e-prints*, page arXiv:2206.14820, juin 2022.
- Warren R. MORNINGSTAR, Yashar D. HEZAVEH, Laurence PERREAULT LEVASSEUR, Roger D. BLANDFORD, Philip J. MARSHALL, Patrick PUTZKY et Risa H. WECHSLER : Analyzing interferometric observations of strong gravitational lenses with recurrent and convolutional neural networks. *arXiv e-prints*, page arXiv:1808.00011, juillet 2018.
- Warren R. MORNINGSTAR, Laurence PERREAULT LEVASSEUR, Yashar D. HEZAVEH, Roger BLANDFORD, Phil MARSHALL, Patrick PUTZKY, Thomas D. RUETER, Risa WECHSLER et Max WELING : Data-driven Reconstruction of Gravitationally Lensed Galaxies Using Recurrent Inference Machines. *ApJ*, 883(1):14, septembre 2019.
- Ethan O. NADLER, Simon BIRRER, Daniel GILMAN, Risa H. WECHSLER, Xiaolong DU, Andrew BENSON, Anna M. NIERENBERG et Tommaso TREU : Dark Matter Constraints from a Unified Analysis of Strong Gravitational Lenses and Milky Way Satellite Galaxies. *ApJ*, 917(1):7, août 2021.
- Julio F. NAVARRO, Carlos S. FRENK et Simon D. M. WHITE : A Universal Density Profile from Hierarchical Clustering. *ApJ*, 490(2):493–508, décembre 1997.

- A. M. NIERENBERG, T. TREU, S. A. WRIGHT, C. D. FASSNACHT et M. W. AUGER : Detection of substructure with adaptive optics integral field spectroscopy of the gravitational lens B1422+231. *MNRAS*, 442(3):2434–2445, août 2014.
- OPENAI : GPT-4 Technical Report. *arXiv e-prints*, page arXiv:2303.08774, mars 2023.
- J. P. OSTRIKER, P. J. E. PEEBLES et A. YAHIL : The Size and Mass of Galaxies, and the Mass of the Universe. *ApJ*, 193:L1, octobre 1974.
- Ji Won PARK, Sebastian WAGNER-CARENA, Simon BIRRER, Philip J. MARSHALL, Joshua Yao-Yu LIN, Aaron ROODMAN et LSST DARK ENERGY SCIENCE COLLABORATION : Large-scale Gravitational Lens Modeling with Bayesian Neural Networks for Accurate and Precise Inference of the Hubble Constant. *ApJ*, 910(1):39, mars 2021.
- Adam PASZKE, Sam GROSS, Francisco MASSA, Adam LERER, James BRADBURY, Gregory CHANAN, Trevor KILLEEN, Zeming LIN, Natalia GIMELSHEIN, Luca ANTIGA *et al.* : Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.
- A. A. PENZIAS : Free Hydrogen in the Pegasus I Cluster of Galaxies. *AJ*, 66:293, mars 1961.
- A. A. PENZIAS et R. W. WILSON : A Measurement of Excess Antenna Temperature at 4080 Mc/s. *ApJ*, 142:419–421, juillet 1965.
- Laurence PERREAULT LEVASSEUR, Yashar D. HEZAVEH et Risa H. WECHSLER : Uncertainties in Parameters Estimated with Neural Networks: Application to Strong Gravitational Lensing. *ApJ*, 850(1):L7, novembre 2017.
- Kaare Brandt PETERSEN, Michael Syskind PEDERSEN *et al.* : The matrix cookbook. *Technical University of Denmark*, 7(15):510, 2008.
- Arlie O. PETERS, Harold LEVINE et Joachim WAMBSGANSS : *Singularity theory and gravitational lensing*. 2001.
- PLANCK COLLABORATION, N. AGHANIM, Y. AKRAMI, M. ASHDOWN, J. AUMONT, C. BACCIGALUPI, M. BALLARDINI, A. J. BANDAY, R. B. BARREIRO, N. BARTOLO, S. BASAK, R. BATTYE, K. BENABED, J. P. BERNARD, M. BERSANELLI, P. BIELEWICZ, J. J. BOCK, J. R. BOND, J. BORRILL, F. R. BOUCHET, F. BOULANGER, M. BUCHER, C. BURIGANA, R. C. BUTLER, E. CALABRESE, J. F. CARDOSO, J. CARRON, A. CHALLINOR, H. C. CHIANG, J. CHLUBA, L. P. L. COLOMBO, C. COMBET, D. CONTRERAS, B. P. CRILL, F. CUTTAIA, P. DE BERNARDIS, G. DE ZOTTI, J. DELABROUILLE, J. M. DELOUIS, E. DI VALENTINO, J. M. DIEGO, O. DORÉ, M. DOUSPIS, A. DUCOUT, X. DUPAC, S. DUSINI, G. EFSTATHIOU, F. ELSNER, T. A. ENSSLIN, H. K. ERIKSEN, Y. FANTAYE, M. FARHANG, J. FERGUSSON, R. FERNANDEZ-COBOS, F. FINELLI, F. FORASTIERI, M. FRAILIS, A. A. FRAISSE, E. FRANCESCHI, A. FROLOV, S. GALEOTTA, S. GALLI,

- K. GANGA, R. T. GÉNOVA-SANTOS, M. GERBINO, T. GHOSH, J. GONZÁLEZ-NUOVO, K. M. GÓRSKI, S. GRATTON, A. GRUPPUSO, J. E. GUDMUNDSSON, J. HAMANN, W. HANDLEY, F. K. HANSEN, D. HERRANZ, S. R. HILDEBRANDT, E. HIVON, Z. HUANG, A. H. JAFFE, W. C. JONES, A. KARAKCI, E. KEIHÄNEN, R. KESKITALO, K. KIIVERI, J. KIM, T. S. KISNER, L. KNOX, N. KRACHMALNICOFF, M. KUNZ, H. KURKI-SUONIO, G. LAGACHE, J. M. LAMARRE, A. LASENBY, M. LATTANZI, C. R. LAWRENCE, M. LE JEUNE, P. LEMOS, J. LESGOURGUES, F. LEVRIER, A. LEWIS, M. LIGUORI, P. B. LILJE, M. LILLEY, V. LINDHOLM, M. LÓPEZ-CANIEGO, P. M. LUBIN, Y. Z. MA, J. F. MACÍAS-PÉREZ, G. MAGGIO, D. MAINO, N. MANDOLESI, A. MANGILLI, A. MARCOS-CABALLERO, M. MARIS, P. G. MARTIN, M. MARTINELLI, E. MARTÍNEZ-GONZÁLEZ, S. MATARRESE, N. MAURI, J. D. McEWEN, P. R. MEINHOLD, A. MELCHIORRI, A. MENNELLA, M. MIGLIACCIO, M. MILLEA, S. MITRA, M. A. MIVILLE-DESCHÊNES, D. MOLINARI, L. MONTIER, G. MORGANTE, A. MOSS, P. NATOLI, H. U. NØRGAARD-NIELSEN, L. PAGANO, D. PAOLETTI, B. PARTRIDGE, G. PATANCHON, H. V. PEIRIS, F. PERROTTA, V. PETTORINO, F. PIACENTINI, L. POLASTRI, G. POLENTA, J. L. PUGET, J. P. RACHEN, M. REINECKE, M. REMAZEILLES, A. RENZI, G. ROCHA, C. ROSSET, G. ROUDIER, J. A. RUBIÑO-MARTÍN, B. RUIZ-GRANADOS, L. SALVATI, M. SANDRI, M. SAVELAINEN, D. SCOTT, E. P. S. SHELLARD, C. SIRIGNANO, G. SIRRI, L. D. SPENCER, R. SUNYAEV, A. S. SUUR-USKI, J. A. TAUBER, D. TAVAGNACCO, M. TENTI, L. TOFFOLATTI, M. TOMASI, T. TROMBETTI, L. VALENZIANO, J. VALIVIITA, B. VAN TENT, L. VIBERT, P. VIELVA, F. VILLA, N. VITTORIO, B. D. WANDEL, I. K. WEHUS, M. WHITE, S. D. M. WHITE, A. ZACCHEI et A. ZONCA : Planck 2018 results. VI. Cosmological parameters. *A&A*, 641:A6, septembre 2020.
- Boris T POLYAK : Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):1–17, 1964.
- Nasim RAHAMAN, Aristide BARATIN, Devansh ARPIT, Felix DRAXLER, Min LIN, Fred A. HAMPRECHT, Yoshua BENGIO et Aaron COURVILLE : On the Spectral Bias of Neural Networks. *arXiv e-prints*, page arXiv:1806.08734, juin 2018.
- A. REFREGIER, A. AMARA, T. D. KITCHING, A. RASSAT, R. SCARAMELLA et J. WELLER : Euclid Imaging Consortium Science Book. *arXiv e-prints*, page arXiv:1001.0061, janvier 2010.
- Adam G. RIESS, Alexei V. FILIPPENKO, Peter CHALLIS, Alejandro CLOCCHIATTI, Alan DIERCKS, Peter M. GARNAVICH, Ron L. GILLILAND, Craig J. HOGAN, Saurabh JHA, Robert P. KIRSHNER, B. LEIBUNDGUT, M. M. PHILLIPS, David REISS, Brian P. SCHMIDT, Robert A. SCHOMMER, R. Chris SMITH, J. SPYROMILIO, Christopher STUBBS, Nicholas B. SUNTZEFF et John TONRY : Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, 116(3):1009–1038, septembre 1998.



- E. RITONDALE, S. VEGETTI, G. DESPALI, M. W. AUGER, L. V. E. KOOPMANS et J. P. MCKEAN : Low-mass halo perturbations in strong gravitational lenses at redshift  $z \sim 0.5$  are consistent with CDM. *MNRAS*, 485(2):2179–2193, mai 2019.
- Frank ROSENBLATT : Principles of neurodynamics. perceptrons and the theory of brain mechanisms. Rapport technique, Cornell Aeronautical Lab Inc Buffalo NY, 1961.
- V. C. RUBIN, Jr. FORD, W. K. et N. THONNARD : Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa  $\rightarrow$  Sc. *ApJ*, 225:L107–L111, novembre 1978.
- V. C. RUBIN, Jr. FORD, W. K. et N. THONNARD : Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R=4kpc) to UGC 2885 (R=122kpc). *ApJ*, 238:471–487, juin 1980.
- Till SAWALA, Carlos S. FRENK, Azadeh FATTAHI, Julio F. NAVARRO, Richard G. BOWER, Robert A. CRAIN, Claudio DALLA VECCHIA, Michelle FURLONG, John. C. HELLY, Adrian JENKINS, Kyle A. OMAN, Matthieu SCHALLER, Joop SCHAYE, Tom THEUNS, James TRAYFORD et Simon D. M. WHITE : The APOSTLE simulations: solutions to the Local Group’s cosmic puzzles. *MNRAS*, 457(2):1931–1943, avril 2016.
- Aurel SCHNEIDER : *Dark matter structures and the free streaming scale*. Thèse de doctorat, University of Zurich, août 2012. URL <https://doi.org/10.5167/uzh-75587>.
- Peter SCHNEIDER : The cosmological lens equation and the equivalent single-plane gravitational lens. *MNRAS*, 292(3):673–678, décembre 1997.
- Peter SCHNEIDER : Generalized multi-plane gravitational lensing: time delays, recursive lens equation, and the mass-sheet transformation. *arXiv e-prints*, page arXiv:1409.0015, août 2014.
- S. SCHULDT, G. CHIRIVÌ, S. H. SUYU, A. YILDIRIM, A. SONNENFELD, A. HALKOLA et G. F. LEWIS : Inner dark matter distribution of the Cosmic Horseshoe (J1148+1930) with gravitational lensing and dynamics. *A&A*, 631:A40, novembre 2019.
- N. SCOVILLE, H. AUSSEL, M. BRUSA, P. CAPAK, C. M. CAROLLO, M. ELVIS, M. GIAVALISCO, L. GUZZO, G. HASINGER, C. IMPEY, J. P. KNEIB, O. LEFEVRE, S. J. LILLY, B. MOBASHER, A. RENZINI, R. M. RICH, D. B. SANDERS, E. SCHINNERER, D. SCHMINOVICH, P. SHOPBELL, Y. TANIGUCHI et N. D. TYSON : The Cosmic Evolution Survey (COSMOS): Overview. *ApJS*, 172(1):1–8, septembre 2007.
- Angelo SECCHI : *L’astronomia in Roma nel pontificato DI Pio IX*. Tipografia della pace, 1877.
- Ravi K. SHETH, H. J. MO et Giuseppe TORMEN : Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes. *MNRAS*, 323(1):1–12, mai 2001.

- Joshua SIMON, Simon BIRRER, Keith BECHTOL, Sukanya CHAKRABARTI, Francis-Yan CYR-RACINE, Ian DELL'ANTONIO, Alex DRLICA-WAGNER, Chris FASSNACHT, Marla GEHA, Daniel GILMAN, Yashar D. HEZAVEH, Dongwon KIM, Ting S. LI, Louis STRIGARI et Tommaso TREU : Testing the Nature of Dark Matter with Extremely Large Telescopes. *BAAS*, 51(3):153, mai 2019.
- Nicolas TESSORE et R. Benton METCALF : The elliptical power law profile lens. *A&A*, 580:A79, août 2015.
- A. VALE et J. P. OSTRICKER : Linking halo mass to galaxy luminosity. *MNRAS*, 353(1):189–200, septembre 2004.
- Ashish VASWANI, Noam SHAZEER, Niki PARMAR, Jakob USZKOREIT, Llion JONES, Aidan N GOMEZ, Łukasz KAISER et Illia POLOSUKHIN : Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- S. VEGETTI, L. V. E. KOOPMANS, M. W. AUGER, T. TREU et A. S. BOLTON : Inference of the cold dark matter substructure mass function at  $z = 0.2$  using strong gravitational lenses. *MNRAS*, 442(3):2017–2035, août 2014.
- S. VEGETTI, L. V. E. KOOPMANS, A. BOLTON, T. TREU et R. GAVAZZI : Detection of a dark substructure through gravitational imaging. *MNRAS*, 408(4):1969–1981, novembre 2010.
- S. VEGETTI, D. J. LAGATTUTA, J. P. MCKEAN, M. W. AUGER, C. D. FASSNACHT et L. V. E. KOOPMANS : Gravitational detection of a low-mass dark satellite galaxy at cosmological distance. *Nature*, 481(7381):341–343, janvier 2012.
- Sebastian WAGNER-CARENA, Jelle AALBERS, Simon BIRRER, Ethan O. NADLER, Elise DARRAGH-FORD, Philip J. MARSHALL et Risa H. WECHSLER : From Images to Dark Matter: End-to-end Inference of Substructure from Hundreds of Strong Gravitational Lenses. *ApJ*, 942(2):75, janvier 2023.
- Sebastian WAGNER-CARENA, Ji Won PARK, Simon BIRRER, Philip J. MARSHALL, Aaron ROODMAN, Risa H. WECHSLER et LSST DARK ENERGY SCIENCE COLLABORATION : Hierarchical Inference with Bayesian Neural Networks: An Application to Strong Gravitational Lensing. *ApJ*, 909(2):187, mars 2021.
- D. WALSH, R. F. CARSWELL et R. J. WEYMANN : 0957+561 A, B: twin quasistellar objects or gravitational lens? *Nature*, 279:381–384, mai 1979.
- D. D. XU, Shude MAO, Andrew P. COOPER, Liang GAO, Carlos S. FRENK, Raul E. ANGULO et John HELLY : On the effects of line-of-sight structures on lensing flux-ratio anomalies in a  $\Lambda$ CDM universe. *MNRAS*, 421(3):2553–2567, avril 2012.

D. D. XU, Shude MAO, Jie WANG, V. SPRINGEL, Liang GAO, S. D. M. WHITE, Carlos S. FRENK, Adrian JENKINS, Guoliang LI et Julio F. NAVARRO : Effects of dark matter substructures on gravitational lensing: results from the Aquarius simulations. *MNRAS*, 398(3):1235–1253, septembre 2009.

F. ZWICKY : On the Masses of Nebulae and of Clusters of Nebulae. *ApJ*, 86:217, octobre 1937.

Ernst ÖPIK : Selective absorption of light in space, and the dynamics of the Universe. *Bulletin de la Société Astronomique de Russie*, 21:150, 1915.



## Annexe A

---

### $\Lambda$ CDM

**Tableau A.1.** Paramètres du modèle  $\Lambda$ CDM mesurés à partir d'observations du fond diffus cosmologique prises par le télescope Planck (Planck Collaboration *et al.*, 2020)

Paramètre	Description	Valeur
$\Omega_{r,0}$	Densité de radiation	$\sim 10^{-4}$
$\Omega_{m,0}$	Densité de matière	0.3111
$\Omega_c h^2$	Densité de matière sombre	0.11933
$\Omega_b h^2$	Densité de matière baryonique	0.02242
$\Omega_{\Lambda,0}$	Densité d'énergie sombre	0.6889
$\Omega_0 h^2$	Densité totale	$\equiv 1$
$h$	Constante de Hubble $h \equiv \frac{H_0}{100 \text{ km s}^{-1} \text{ Mpc}^{-1}}$	0.6766