The hunt for stellar-mass DM clumps: applying the statistical machine learning techniques to strong microlensing events

Owerview of my talk:

- Gravitational microlensing (GM): some basics
- Dark matter clumps in SIDM and CDM
- High amplification events (HAE): DM vs. point lens photometrical appearances
- Fitting the HAE lightcurves (“standard way”)
- SML vs. “standard fitting”
- ...

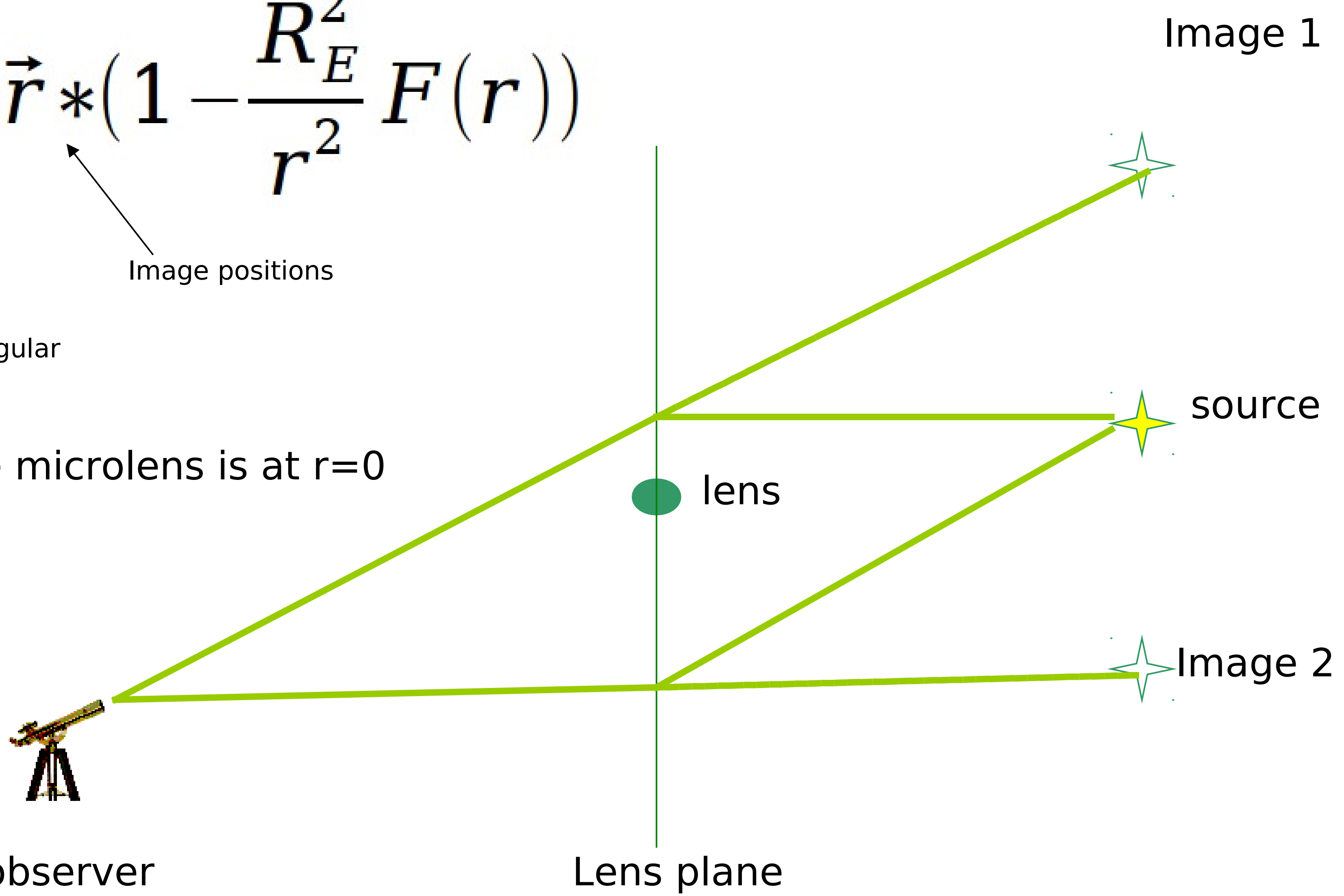
Some basics of gravitational (micro)lensing

$$\vec{\eta} = \vec{r} * \left(1 - \frac{R_E^2}{r^2} F(r)\right)$$

Image positions

Source angular position

Single microlens is at r=0



Gravitational lensing hierarchy and observational appearances

class	Lensing objects	Observational appearances
Macrolensing	Galaxies, > 10^9 M_{Sun}	Astrometric (static multiple images) Photometric (time delays between images)
Mesolensing	Globular clusters, DM substructure 10^4-10^8 M_{Sun}	Photometric (slow changes, >monthly-yearly timescales, anomalous flux ratios) Astrometric (~milliarcsec image splitting)
Microlensing	Stellar-mass objects, 10^{-1}-10^3 M_{Sun}	Photometric (caustic crossing high amplification events, daily-weekly timescales) Spectral (emission lines profiles distortions)
Nanolensing	Planetary mass objects, <10^{-2} $M_{\odot}$	Photometric, complementary to microlensing

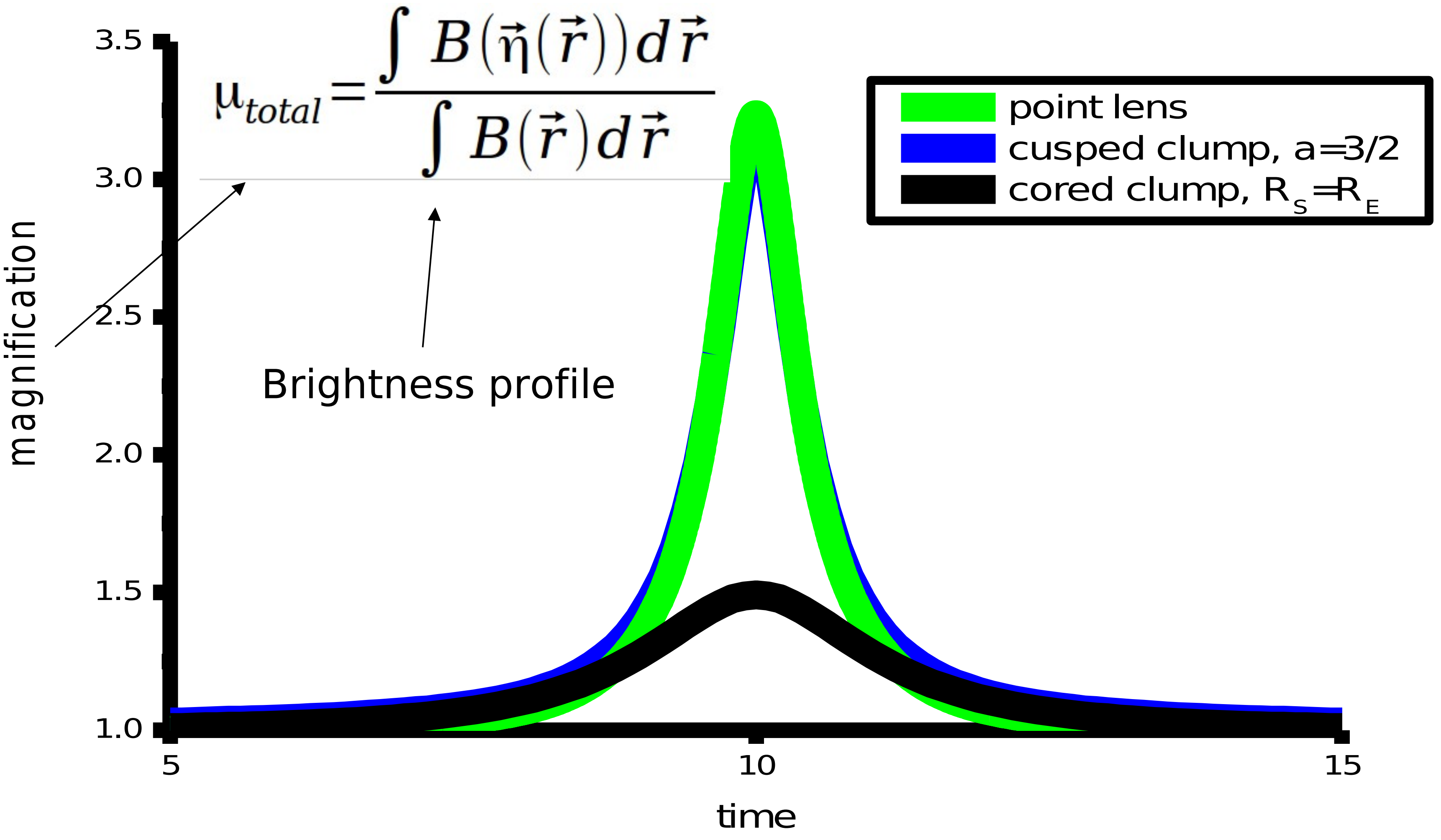
DM clumps as microlenses

$$\vec{\eta} = \vec{r} * \left(1 - \frac{R_E^2}{r^2} F(r)\right)$$

DM model	(micro)lensing equation
Cold (CDM) cusped clump	$F(r) = (r/R_c)^a$, $a<2$
Self-interacting (SIDM) cored clump	$F(r) = r^2/(r^2+R_s^2)$
Warm DM, no stellar-mass clumps	$F(r) = 0$
Ultra compact mini halo (UCMH)	$F(r) = -1$ (i.e. a star surrounded with a disrupted DM clump)

For details: Gravitational microlensing as a probe for dark matter clumps, MNRAS 2015

High amplification events



- Standard fitting of one ligtcurve consumes 10 minutes - 1 hour

SML parameters determining: correlations

algo-rithm	Impact param.	Power index a	Lens size	Source size	sign	Processing time, sec
Linear regress.	0.45	0.94	0.65	0.62	0.88	~1
KNN-3	0.87	0.96	0.96	0.98	0.98	~0.25
KNN-5	0.85	0.96	0.95	0.98	0.98	~0.1
Random tree	0.83	0.96	0.93	0.98	0.98	~0.25
Random forest	0.83	0.96	0.98	0.98	0.99	~0.25
Decision table	0.82	0.94	0.91	0.96	0.99	~1
Multil. percept.	0.84	0.96	0.96	0.98	0.97	~35

SML vs. standard fitting

- Amongst the algorithms considered, Random forest show the best speed/quality (~0.25 sec for the set of 300 curves training + 300 curves testing);
- … and despite for bigger sets + observational data (like MACHOs) with error bars, this process will be slower and with quality a bit lower…
- … it is far less time-consuming that “standard way” fitting and allows us to analyze many curves simultaneously.
- Thank You for Your attention!