INITIATION À LA RECHERCHE

Master 1 IDSM

Iuliia Tkachenko

iuliia.tkachenko@univ-lyon2.fr

L'ÉDITION SCIENTIFIQUE AVEC LATEX

Le mode mathématique (1/2)

```
\usepackage{amsmath,amssymb}
```

- Attention, les packages `amssymb` et `wasysym` sont incompatibles.
- Passage en mode mathématique avec `$`

```
On a $3x+1=y$ où $y < 1$.
```

On a $3x+1=y$ où $y < 1$.

```
$\text{On a } 3x+1=y \text{ où } y<1.$
```

```
Notons $f$ la fonction.
```

Notons f la fonction.

Le mode mathématique (2/2)

- Mode mathématique centré avec `$$`
- Indices et exposants :

```
$x_i = x^{3a+b}$
```

$$x_i = x^{3a+b}$$

```
$x_i^n \neq {x_i}^n$
```

$$x_i^n \neq {x_i}^n$$

Fractions, racines et fonctions

$\frac{a}{b}$	`$\frac{a}{b}$`	`$\tfrac{a}{b}$` ou	`$$\tfrac{a}{b}$$`
$\frac{a}{b}$	`$$\frac{a}{b}$$`	`$\dfrac{a}{b}$` ou	`$$\dfrac{a}{b}$$`

```
$\sqrt{4}=\sqrt[3]{8}$
```

$\sqrt{4}=\sqrt[3]{8}$

lim	`\lim`	Pr	`\Pr`	$\overline{\lim}$	`\varlimsup`	det	`\det`
lim inf	`\liminf`	inf	`\inf`	$\underline{\lim}$	`\varliminf`	max	`\max`
lim sup	`\limsup`	sup	`\sup`	gcd	`\gcd`	min	`\min`
cos	`\cos`	cot	`\cot`	exp	`\exp`	hom	`\hom`
sin	`\sin`	cosh	`\cosh`	ln	`\ln`	dim	`\dim`
tan	`\tan`	sinh	`\sinh`	log	`\log`	ker	`\ker`
arccos	`\arccos`	tanh	`\tanh`	deg	`\deg`	csc	`\csc`
arcsin	`\arcsin`	coth	`\coth`	(mod q)	`\pmod q`	lg	`\lg`
arctan	`\arctan`	arg	`\arg`	mod q	`\mod q`	sec	`\sec`

Disposition des indices et des exposants, sommes, intégrales et produits

$\lim_{x \to 0}$ `$\lim_{x \to 0}$`
ou `$$\lim\nolimits_{x \to 0}$$`

$$\lim_{x \to 0}$$ `$$\lim_{x \to 0}$$`
ou `$\lim\limits_{x \to 0}$`

$\int$ `\int`	$\iint$ `\iint`	$\iiint$ `\iiint`
$\oint$ `\oint`	$\iiiint$ `\iiiint`	$\int \cdots \int$ `\idotsint`
$\sum$ `\sum`	$\prod$ `\prod`	$\coprod$ `\coprod`

Les espaces

- Gestion automatique des espaces en mode mathématique :

`$a=3+b$` a=3+b

`$a= 3 + b$` a=3+b

Type d'espace	commande	*AA*	valeur (cadratins)
négatif	`\!`	*AA*	$-3/18$
fin	`\,`	*A A*	$1/18$
moyen	`\:`	*A A*	$3/18$
large	`\;`	*A A*	$4/18$
blanc normal	`\␣`	*A A*	*(variable)*
cadratin	`\quad`	*A A*	1
double cadratin	`\qquad`	*A A*	2

Signes, chapeaux et accents

$\hat{a}$	\hat{a}	$\dot{a}$	\dot{a}	$\tilde{a}$	\tilde{a}
$\bar{a}$	\bar{a}	$\ddot{a}$	\ddot{a}	$\check{a}$	\check{a}
$\vec{a}$	\vec{a}	$\dddot{a}$	\dddot{a}	$\breve{a}$	\breve{a}
$\acute{a}$	\acute{a}	$\ddddot{a}$	\ddddot{a}		
$\grave{a}$	\grave{a}	$\mathring{a}$	\mathring{a}		

$\vec{\imath}, \vec{\jmath}$ (et non $\vec{i}$) `$\vec{\imath}, \vec{\jmath}$`

$\widetilde{AB}$	\widetilde{AB}	$\widehat{AB}$	\widehat{AB}
$\underline{AB}$	\underline{AB}	$\overline{AB}$	\overline{AB}
$\overrightarrow{AB}$	\overrightarrow{AB}		

Symboles classiques (1/2)

∞ \infty	$\exists$ \exists	$\varnothing$ \varnothing	$\hbar$ \hslash
ℓ \ell	$\forall$ \forall	$\imath$ \imath	$\hbar$ \hbar
$\Im$ \Im	∇ \nabla	$\jmath$ \jmath	$\wp$ \wp
$\Re$ \Re	∂ \partial	$\aleph$ \aleph	$\top$ \top
$\flat$ \flat	$\natural$ \natural	$\sharp$ \sharp	$\bot$ \bot
$\ll$ \ll	$\approx$ \approx	$\parallel$ \parallel	$\lhd$ \lhd
$\gg$ \gg	$\leqslant$ \leqslant	$\subset$ \subset	$\in$ \in
$\equiv$ \equiv	$\geqslant$ \geqslant	$\supset$ \supset	$\ni$ \ni
$\sim$ \sim	$\propto$ \propto	$\subseteq$ \subseteq	$\mid$ \mid
$\simeq$ \simeq	$\perp$ \perp	$\varsubsetneq$ \varsubsetneq	$\neg$ \neg

Plus d'info : https://tug.ctan.org/info/symbols/comprehensive/symbols-a4.pdf

Recherche d'un symbole en ligne : http://detexify.kirelabs.org/classify.html

Symboles classiques (2/2)

$\pm$ \pm	$\bigcirc$ \bigcirc	$\circledast$ \circledast	$\setminus$ \setminus
$\mp$ \mp	$\Diamond$ \Diamond	$\boxdot$ \boxdot	$\cap$ \cap
$*$ \ast	$\bullet$ \bullet	$\boxplus$ \boxplus	$\cup$ \cup
$\star$ \star	$\odot$ \odot	$\boxminus$ \boxminus	$\rtimes$ \rtimes
$\times$ \times	$\oplus$ \oplus	$\boxtimes$ \boxtimes	$\ltimes$ \ltimes
$\uplus$ \uplus	$\ominus$ \ominus	$\Box$ \Box	$\vee$ \vee
$\sqcup$ \sqcup	$\oslash$ \oslash	$\complement$ \complement	$\wedge$ \wedge
$\circ$ \circ	$\otimes$ \otimes	$\smallsetminus$ \smallsetminus	$\models$ \models
$\bigcap$ \bigcap	$\bigwedge$ \bigwedge	$\bigotimes$ \bigotimes	
$\bigcup$ \bigcup	$\biguplus$ \biguplus	$\bigoplus$ \bigoplus	
$\bigvee$ \bigvee	$\bigsqcup$ \bigsqcup	$\bigodot$ \bigodot	
$/$ \diagup	$\diagdown$ \diagdown	$\backslash$ \backslash	

Lettres grecques

α \alpha	θ \theta	π \pi	ϕ \phi
β \beta	ϑ \vartheta	ϖ \varpi	φ \varphi
γ \gamma	ι \iota	ρ \rho	χ \chi
δ \delta	κ \kappa	ϱ \varrho	ψ \psi
ϵ \epsilon	λ \lambda	σ \sigma	ω \omega
ε \varepsilon	μ \mu	ς \varsigma	
ζ \zeta	ν \nu	τ \tau	
η \eta	ξ \xi	υ \upsilon	
Γ \Gamma	Λ \Lambda	Σ \Sigma	Ψ \Psi
Δ \Delta	Ξ \Xi	Υ \Upsilon	Ω \Omega
Θ \Theta	Π \Pi	Φ \Phi	

Flèches

\leftarrow donne $\leftarrow$ et \Downarrow permet d'obtenir $\Downarrow$.

$\rightarrow$ \rightarrow (synonyme : \to)	$\hookrightarrow$ \hookrightarrow
$\longrightarrow$ \longrightarrow	$\rightharpoonup$ \rightharpoonup
$\Rightarrow$ \Rightarrow	$\circlearrowright$ \circlearrowright
$\Longrightarrow$ \Longrightarrow	$\curvearrowright$ \curvearrowright
$\dashrightarrow$ \dashrightarrow	$\uparrow$ \uparrow
$\rightrightarrows$ \rightrightarrows	$\Uparrow$ \Uparrow
$\twoheadrightarrow$ \twoheadrightarrow	
$\leftrightarrow$ \leftrightarrow	$\mapsto$ \mapsto
$\longleftrightarrow$ \longleftrightarrow	$\longmapsto$ \longmapsto
$\Leftrightarrow$ \Leftrightarrow	$\nearrow$ \nearrow
$\Longleftrightarrow$ \Longleftrightarrow (syn. : \iff)	$\nwarrow$ \nwarrow
$\leftrightarrows$ \leftrightarrows	$\searrow$ \searrow
$\rightleftarrows$ \rightleftarrows	$\swarrow$ \swarrow
$\leftrightharpoons$ \leftrightharpoons	$\updownarrow$ \updownarrow
$\rightleftharpoons$ \rightleftharpoons	$\Updownarrow$ \Updownarrow
$\leadsto$ \leadsto	

Négations des symboles relationnels

Faire précéder la commande de symbole relationnel de `\not`

```
$A \not\subset E$
```

$A \not\subset E$

$\neq$ \neq	$\nmid$ \nmid	$\nRightarrow$ \nRightarrow
$\nsim$ \nsim	$\nparallel$ \nparallel	$\nLeftarrow$ \nLeftarrow
$\nexists$ \nexists	$\nrightarrow$ \nrightarrow	$\nleftrightarrow$ \nleftrightarrow
$\notin$ \notin	$\nleftarrow$ \nleftarrow	$\nLeftrightarrow$ \nLeftrightarrow

Parenthèses extensibles (1/2)

```
$\left( \dfrac{a}{b} \right)$
```
$\left(\dfrac{a}{b} \right)$

```
$\left|\frac{\phi(t)}{3}
\right\rangle$
```
$\left|\frac{\phi(t)}{3}\right\rangle$

```
$\left. \dfrac{\partial f}
{\partial T} \right)_{P,V}$
```
$\left. \dfrac{\partial f}{\partial T} \right)_{P,V}$

Cas particulier :

```
$\left( \overbrace{AB^2 +
BC^2}^{\text{Pythagore}}
\right)$
```
$\left(\overbrace{AB^2 + BC^2}^{\text{Pythagore}} \right)$

\big, \Big, \bigg et \Bigg (par ordre croissant)

```
$\big( \overbrace{AB^2 +
BC^2}^{\text{Pythagore}}
\big)$
```
$\big(\overbrace{AB^2 + BC^2}^{\text{Pythagore}} \big)$

Parenthèses extensibles (2/2)

(	(	{	\{	⟨	\langle	↕	\updownarrow
)	)	}	\}	⟩	\rangle	⇕	\Updownarrow
[	[	⌊	\lfloor	↑	\uparrow	\	\backslash
]	]	⌋	\rfloor	⇑	\Uparrow	‖	|
/	/	⌈	\lceil	↓	\downarrow	⟦	\llbracket†
|	|	⌉	\rceil	⇓	\Downarrow	⟧	\rrbracket†

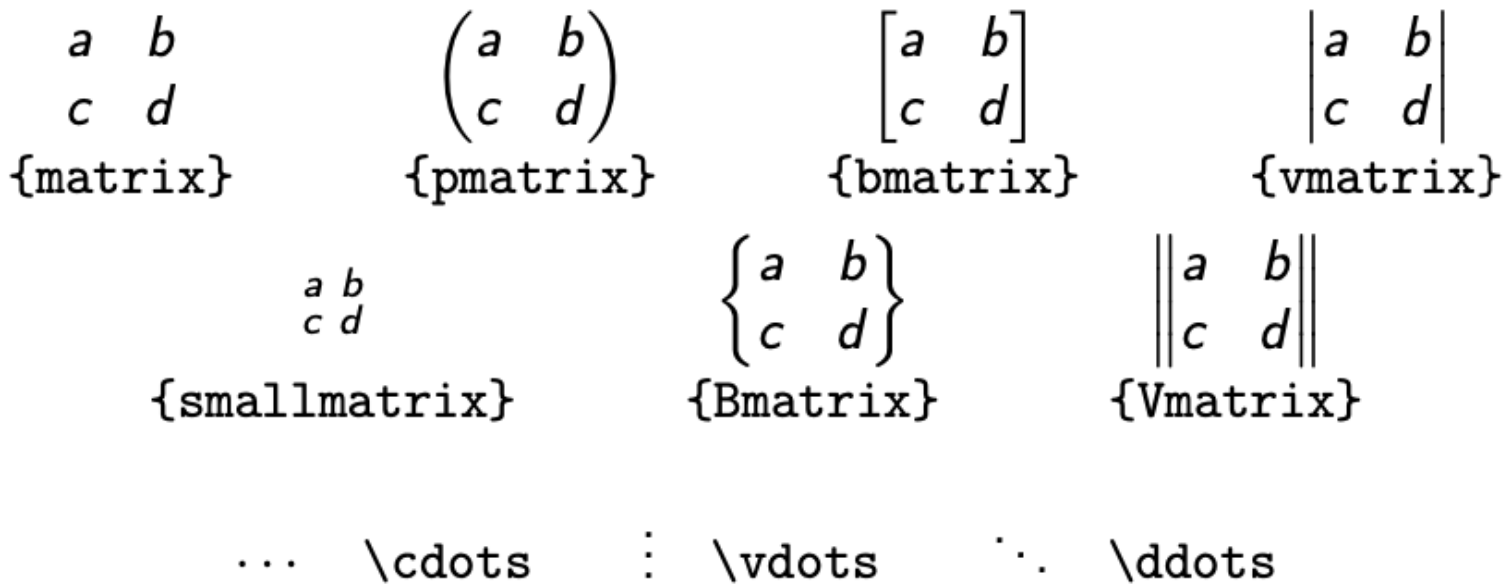
† commandes du package stmaryrd

Tableaux

$f(t)$	$F(p)$
1	$1/p$
t	$1/p^2$

```
$\begin{array}{|c|c|}
    \hline     f(t) & F(p)  \\[1mm]
    \hline \hline 1 & 1/p   \\[1mm]
                  t & 1/p^2 \\ \hline
 \end{array}$
```

Matrices (1/2)


a b
c d
{matrix}
{pmatrix}
{bmatrix}
{vmatrix}
{smallmatrix}
{Bmatrix}
{Vmatrix}
\cdots
\cdots
\vdots
\vdots
\ddots
\ddots

Matrices (2/2)

```
$\begin{pmatrix}
  \cos\theta & -\sin\theta \\
  \sin\theta & \cos\theta
\end{pmatrix}$
```

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}$$

```
$\begin{pmatrix}
1 & \phantom{-}1 \\
0 & -1 \end{pmatrix}$
```

Numérotation des équations

```
\begin{equation}
  y'' -\omega^2 y = f
  \label{eq:ED1}
\end{equation}
L'équation~\eqref{eq:ED1} implique la continuité de~$y$.
```

$$y'' - \omega^2 y = f \tag{1}$$

L'équation (1) implique la continuité de y.

Pour redémarrer la numérotation à chaque section par exemple :

```
\numberwithin{equation}{section}
```

Structures conditionnelles

$$\delta_{ij} = \begin{cases} 0 & \text{si } i \neq j \\ 1 & \text{si } i = j \end{cases}$$

```
\[ \delta_{ij} =
\begin{cases}
  0 & \text{si } i\neq j\\
  1 &\text{si } i=j
\end{cases} \]
```

Images multiples (1/2)

```
\begin{minipage}[b]{0.45\linewidth}
  \centering
  \centerline{\includegraphics[scale=0.05]{logo_lyon2.png}}
  \bigskip
  \centerline{\small{(a)}}
\end{minipage}
\hfill
\begin{minipage}[b]{.45\linewidth}
  \centering
  \centerline{\includegraphics[scale=0.25]{logo_icom.png}}
  \bigskip
  \centerline{\small{(b)}}
 \end{minipage}
```

(a)

(b)

Images multiples (2/2)

```
\begin{figure}[htb]
\begin{minipage}[b]{0.45\linewidth}
  \centering
  \centerline{\includegraphics[scale=0.05]{logo_lyon2.png}}
  \bigskip
  \centerline{\small{(a)}}
\end{minipage}
\hfill
\begin{minipage}[b]{.45\linewidth}
  \centering
  \centerline{\includegraphics[scale=0.25]{logo_icom.png}}
  \bigskip
  \centerline{\small{(b)}}
 \end{minipage}
 \centering
\caption{Logo a) Université Lyon2 and b) Institut de la Communication.}
\label{fig:logos}
\end{figure}
```

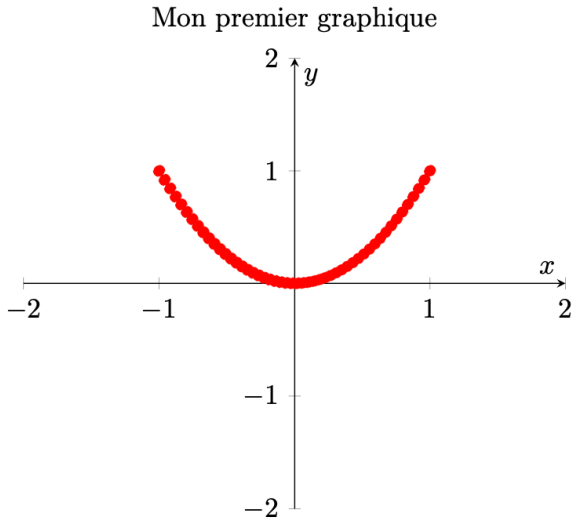
(a)

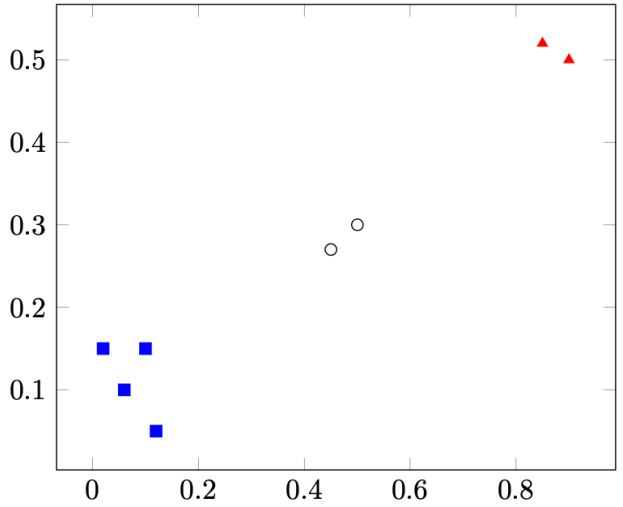
(b)

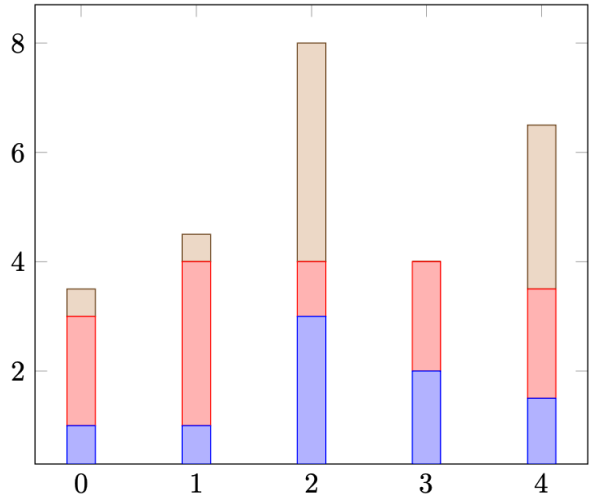
FIGURE 1 – Logo a) Université Lyon2 and b) Institut de la Communication.

Graphiques avec LATEX

```
\usepackage{pgfplots}
```



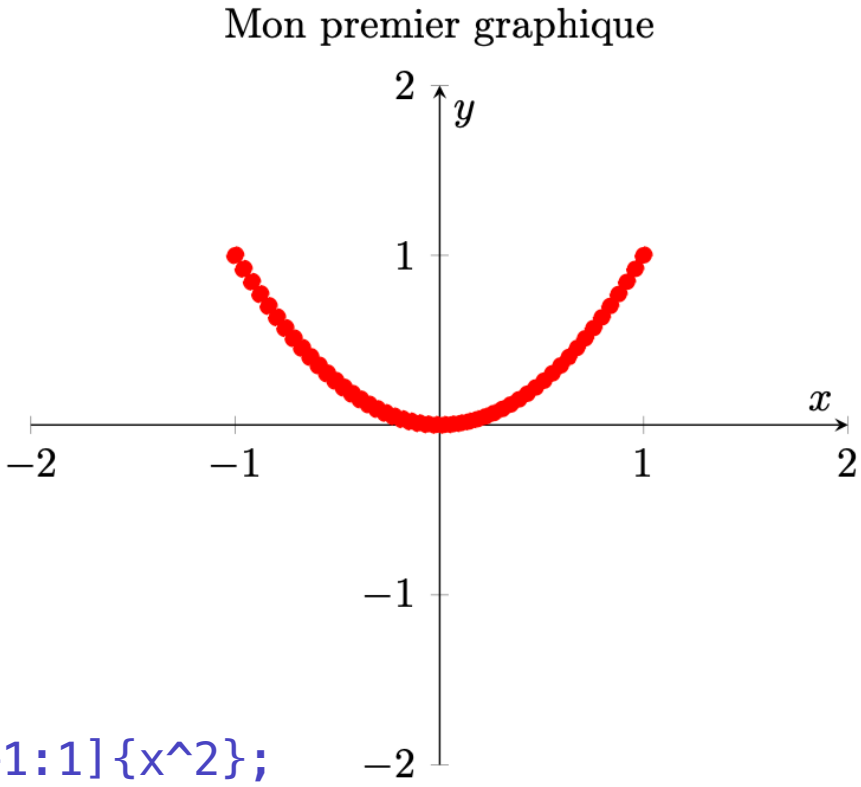




Les graphiques inspirés du tutoriel : https://www.youtube.com/watch?v=5jmIHOWpEg0&t=1456s

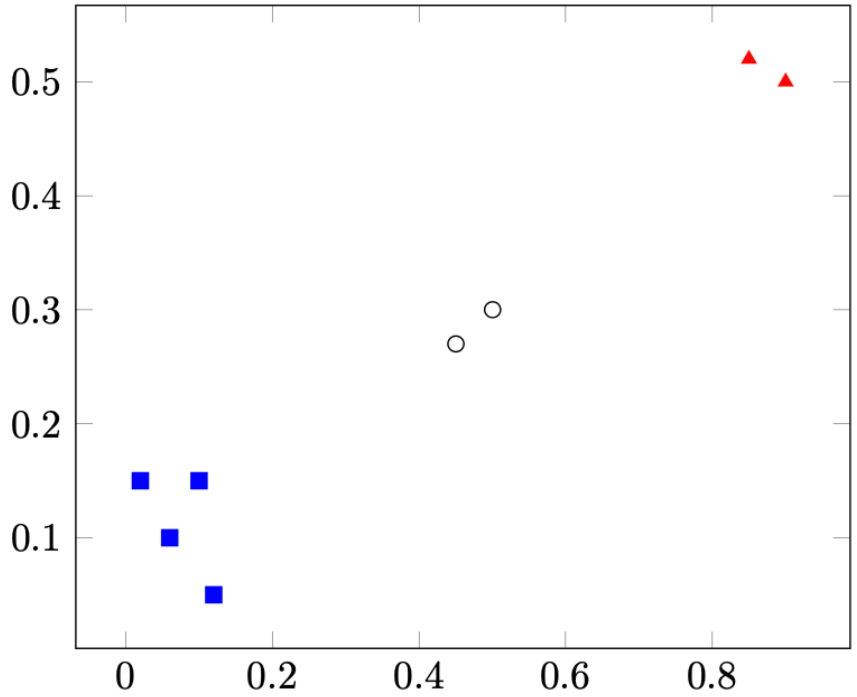
Dessiner une courbe

```
\begin{tikzpicture}
\begin{axis}[xmin=-2, xmax=2, ymin=-2, ymax=2,
axis lines = middle,
xlabel=$x$,
ylabel=$y$,
title={Mon premier graphique}]
\addplot[color=red, dashed, mark=*, samples =50, domain=-1:1]{x^2};
\end{axis}
\end{tikzpicture}
```



Graphique (1/2)

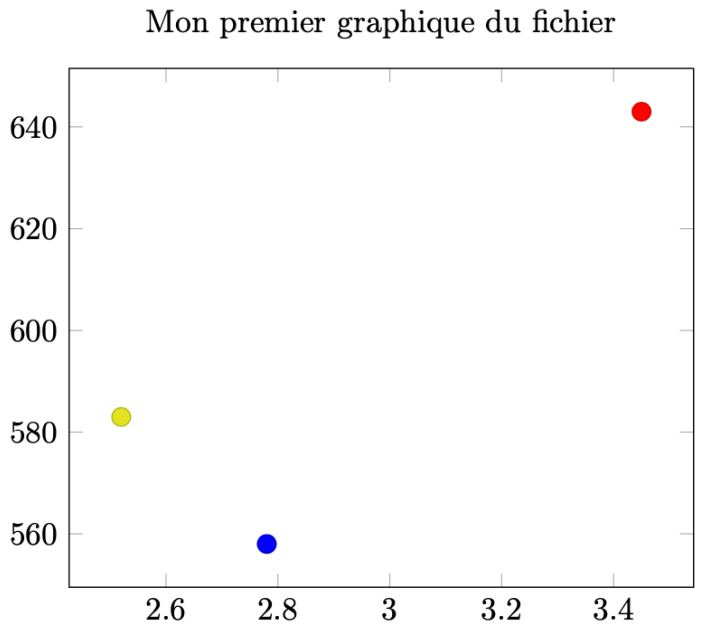
```
\begin{tikzpicture}
  \begin{axis}[scatter/classes={ a={mark=square*,blue}, b={mark=triangle*,red}, c={mark=o,draw=black}}]
   \addplot[scatter, only marks, scatter src=explicit symbolic]
       coordinates {
           (0.1,0.15)[a]
           (0.45,0.27)[c]
           (0.02,0.15)[a]
           (0.06,0.1)[a]
           (0.9,0.5)[b]
           (0.5,0.3)[c]
           (0.85,0.52)[b]
           (0.12,0.05)[a]
       };
  \end{axis}
\end{tikzpicture}
```



Graphique (2/2)

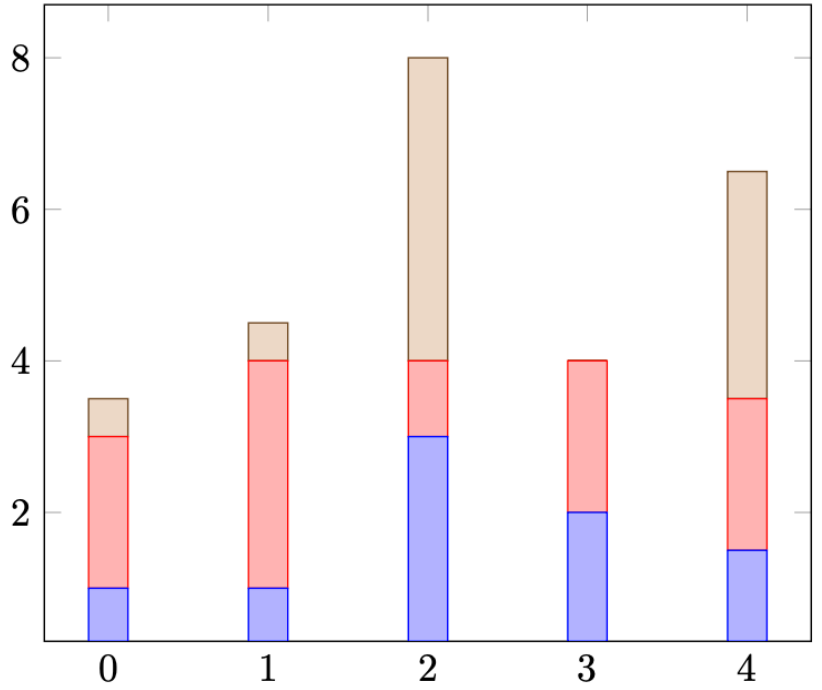
- Utilisation des données d’un fichier .txt

```
\begin{tikzpicture}
  \begin{axis} [title={Mon premier graphique du fichier}]
  \addplot+[only marks, scatter, mark size=2.9pt]
  table[meta=ma] %quelles données on utilise comme axe y
  {test.txt};
  \end{axis}
\end{tikzpicture}
```



Histogramme

```
\begin{tikzpicture}
  \begin{axis} [ybar stacked]
  \addplot coordinates
  {(0,1) (1,1) (2,3) (3,2) (4,1.5)};
  \addplot coordinates
  {(0,2) (1,3) (2,1) (3,2) (4,2)};
  \addplot coordinates
  {(0,0.5) (1,0.5) (2,4) (3,0) (4,3)};
  \end{axis}
\end{tikzpicture}
```



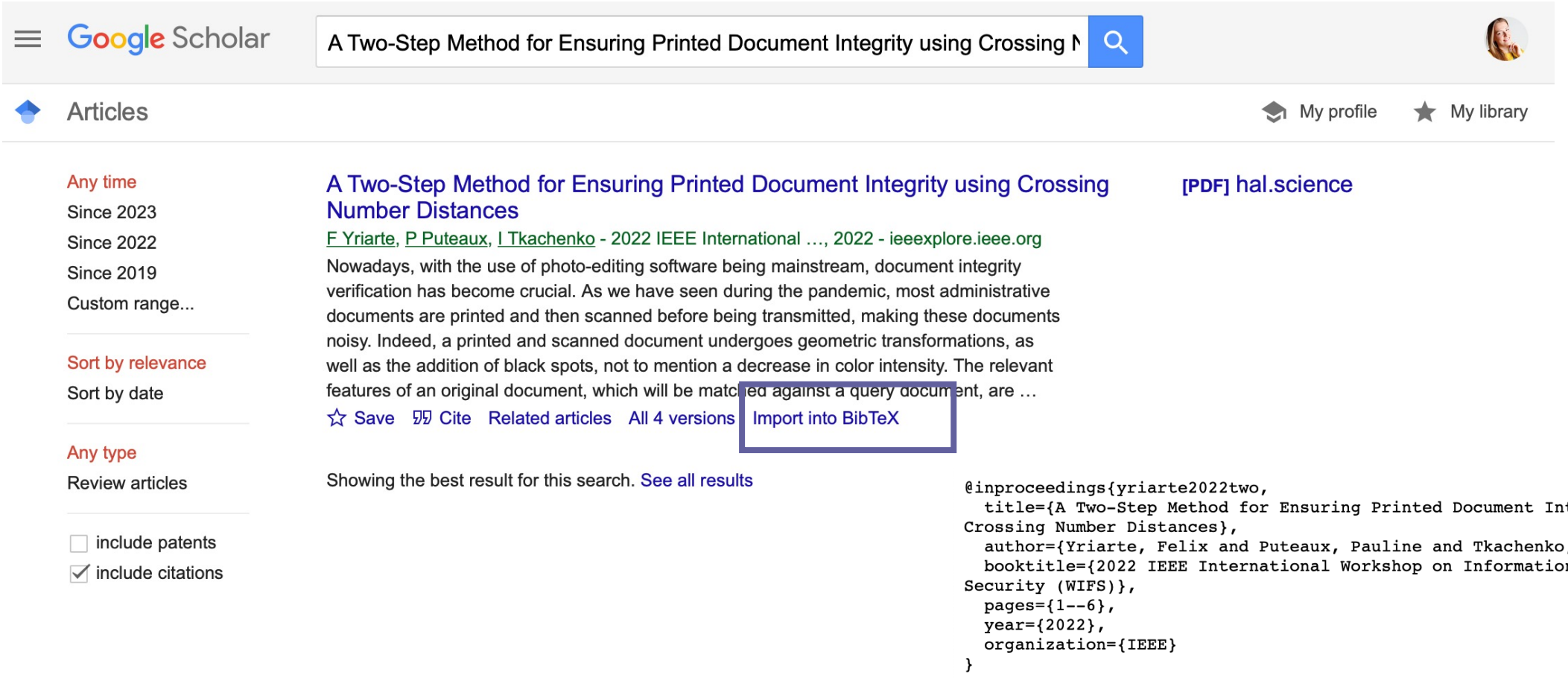
Utiliser la classe d’une conférence ou d’un journal

- Template IEEE https://www.ieee.org/conferences/publishing/templates.html
- Template Elsevier https://www.elsevier.com/authors/policies-and-guidelines/latex-instructions

La bibliographie : BibTEX

- Vous devez stocker les informations dans un fichier dont l'extension est `.bib`, disons `ref.bib` .
- Vous pouvez garder le même fichier pour tous vos documents : BibTEX n'y prend que ce dont il a besoin.
- Outils de création :
  - Editeur d'une publication
  - Google Scholar
  - plugin Zotero de Firefox
  - Jabref

Google Scholar



La bibliographie : syntaxe

Voici la syntaxe :

```
@BOOK{mathaz,
                author = "Hauchecorne, Bertrand AND Suratteau, Daniel",
                title = "Des Mathématiciens de A à Z",
                publisher = "Ellipses",
                year = "1999"
}
```

Citation : `\cite{mathaz}` ou `\cite[p.~203]{mathaz}`

Apparition dans la biblio sans citation : `\nocite{mathaz}`

On fait appel au programme externe `bibtex`.

La bibliographie : section

```
\bibliographystyle{plain}
\bibliography{ref}
```

`plain` : trie les entrées par ordre alphabétique d'auteur et les numérote séquentiellement entre crochets.

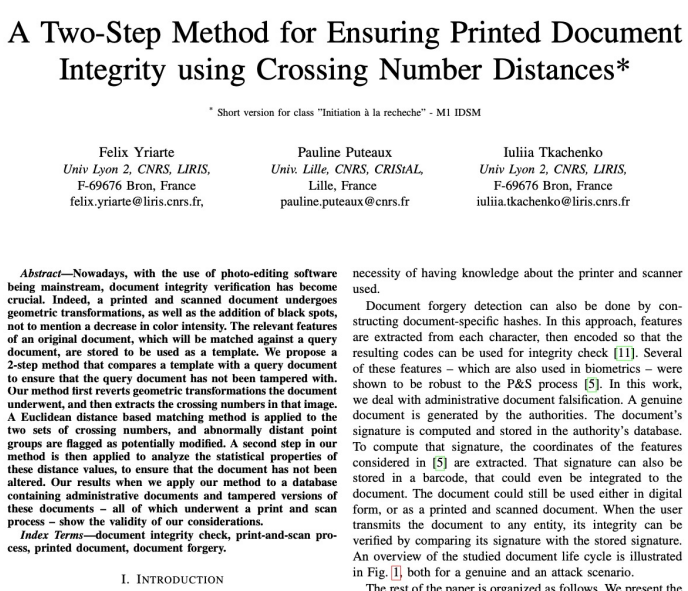
`abbrv` : identique à plain, en abrégeant certains champs comme les prénoms, les mois ou les noms des revues.

`unsrt` : trie les entrées par ordre d'apparition relevé lors du premier passage de LATEX et les numérote séquentiellement entre crochets.

`alpha` : trie les entrées par ordre alphabétique d'auteur et repère les occurrences par les trois premières lettres du nom de l'auteur suivi des deux derniers chiffres de l'année. Un texte de Napoléon datant de 1805 sera noté [Nap05].

Exercice

- Utilisez un template d'article IEEE sur Overleaf
- Re-editez le papier « A Two-Step Method for Ensuring Printed Document Integrity using Crossing Number Distances », 4 pages
- Le papier et les images nécessaires sont en ligne sur la page du cours



A Two-Step Method for Ensuring Printed Document Integrity using Crossing Number Distances*

* Short version for class "Initiation à la recheche" - M1 IDSM

Felix Yriarte
Univ Lyon 2, CNRS, LIRIS,
F-69676 Bron, France
felix.yriarte@liris.cnrs.fr,

Pauline Puteaux
Univ. Lille, CNRS, CRIStAL,
Lille, France
pauline.puteaux@cnrs.fr

Iuliia Tkachenko
Univ Lyon 2, CNRS, LIRIS,
F-69676 Bron, France
iuliia.tkachenko@liris.cnrs.fr

***Abstract*—Nowadays, with the use of photo-editing software being mainstream, document integrity verification has become crucial. Indeed, a printed and scanned document undergoes geometric transformations, as well as the addition of black spots, not to mention a decrease in color intensity. The relevant features of an original document, which will be matched against a query document, are stored to be used as a template. We propose a 2-step method that compares a template with a query document to ensure that the query document has not been tampered with. Our method first reverts geometric transformations the document underwent, and then extracts the crossing numbers in that image. A Euclidean distance based matching method is applied to the two sets of crossing numbers, and abnormally distant point groups are flagged as potentially modified. A second step in our method is then applied to analyze the statistical properties of these distance values, to ensure that the document has not been altered. Our results when we apply our method to a database containing administrative documents and tampered versions of these documents – all of which underwent a print and scan process – show the validity of our considerations.**

***Index Terms*—document integrity check, print-and-scan process, printed document, document forgery.**

I. Introduction

necessity of having knowledge about the printer and scanner used.

Document forgery detection can also be done by constructing document-specific hashes. In this approach, features are extracted from each character, then encoded so that the resulting codes can be used for integrity check [11]. Several of these features – which are also used in biometrics – were shown to be robust to the P&S process [5]. In this work, we deal with administrative document falsification. A genuine document is generated by the authorities. The document's signature is computed and stored in the authority's database. To compute that signature, the coordinates of the features considered in [5] are extracted. That signature can also be stored in a barcode, that could even be integrated to the document. The document could still be used either in digital form, or as a printed and scanned document. When the user transmits the document to any entity, its integrity can be verified by comparing its signature with the stored signature. An overview of the studied document life cycle is illustrated in Fig. 1, both for a genuine and an attack scenario.

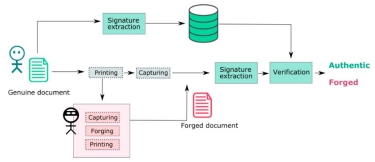


Fig. 1. Overview of the considered scenario: the green blocks correspond to authorities, the red blocks correspond to a forger. The dashed parts represent optional processes.

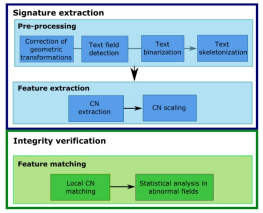


Fig. 2. Pipeline of the document image pre-processing steps and text integrity verification.

A. Pre-processing operations

The P&S process adds different types of degradations to a document: geometric transformations, addition of black spots, gray-level value change (P&S images are grayscale, whereas digital images usually are binary), as well as compression artifacts that appear after the scan [9]. In order to eliminate part of the deterioration that is due to the P&S process, several pre-processing operations are applied to a document image. Here we list the pre-processing operations that are shown in our document integrity check pipeline illustrated in Fig. 2.

- **Correction of geometric transformations.** Scanning a printed document adds different geometric transformations to the document image, such as rotation, translation, scale change, as well as crop. Even if the scanning operation is done carefully, the resulting image is bound to undergo geometric transformations.
- **Text content detection.** We then apply a text detection step so as to remove spots, and non-textual image contents, such as tables, or the document frame. Indeed, after a P&S process, it is common to find small black spots on the document image. These spots usually are smaller than a character.
  We apply the edge-detection method proposed by Suzuki and Keiichi [10], and filter out unusually small or large content. This allows us to shape an image mask – a binary image of the same resolution as the document image – indicating the fields that are to be considered.
- **Binarization and skeletonization.** In order to extract the feature points from a document image, it is necessary to apply a skeletonization step to it. A skeletonization is a shape thinning operation that returns a 1 pixel-wide skeleton that preserves the connectivity of components. We use the method proposed by Lee in [2]. Before the skeletonization, the image is thresholded using the thresholding method proposed by Otsu [3].

After the application of these pre-processing operations, a skeleton image is obtained, and features can be extracted.

B. Feature extraction

Since the P&S process induces noise, it is important to take into account features that are robust to such operations. Crossing Numbers (CN) are used in biometrics (namely, fingerprint recognition) as features because of their stability through acquisition noise. They can however also be used to identify textual characters [5], [11]. A skeleton pixel's crossing number value is the number of neighbouring skeleton pixel it has. This gives an indication on the connectivity of every skeleton pixel. We do not consider CN values of 2 for our matching step, since they cannot be used to distinguish different characters.

Furthermore, some characters can present serifs – small lines or strokes attached to the end of longer strokes. Serifs cause the addition of two types of crossing numbers: a bifurcation and up to two ending points. At lower resolutions, serifs are not reliably identified and can be missed ; in which case the matching step fails for CN feature points extracted from serifs. Since these features are not significant for the matching step, we remove them from the feature point set during the CN feature point extraction. Depending on the scanning quality used, the resulting P&S image's resolution differs. It is thus necessary to scale the extracted CN feature points so as to normalize the data. However, we do not want to induce interpolation errors by applying a scaling operation to the image, which could add discontinuities to a skeleton image. That is why we scale the coordinates of the extracted CN feature points, rather than scaling the document image itself. To do so, every extracted CN feature point's coordinates are multiplied by the ratio $\frac{\text{Reference\_Image\_Resolution}}{\text{Query\_Image\_Resolution}}$. Note that every reference document – since they are numeric – has the same image resolution. We consider that resolution known when scaling the extracted CN feature points.

III. Document integrity check

After the CN feature points are extracted from the query document image, they are matched against the reference

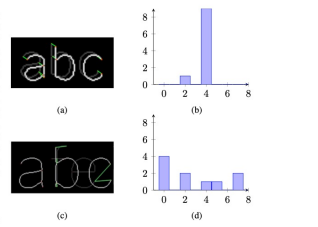


Fig. 3. Example of a) identical shifted sequences of characters (authentic