

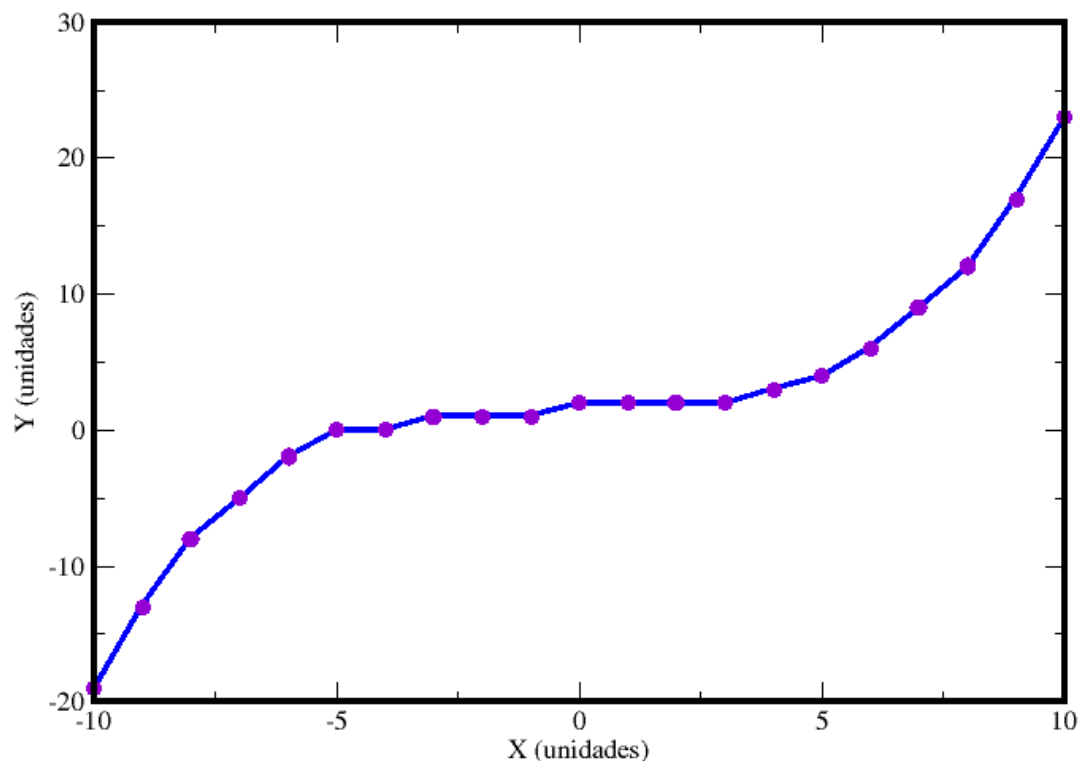
GRÁFICOS E REGRESSÃO LINEAR

J.R. Kaschny
(IFBA 2023)

Introdução

Em termos rudimentares, um gráfico (bidimensional) é basicamente uma forma elegante e eficiente de observarmos o comportamento de uma dada grandeza (eixo vertical) quando expressa como função de uma outra grandeza (eixo horizontal). Como podemos observar abaixo, um gráfico se mostra mais confortável que uma tabulação de dados, permitindo uma interpretação visual mais eficiente.^(*)

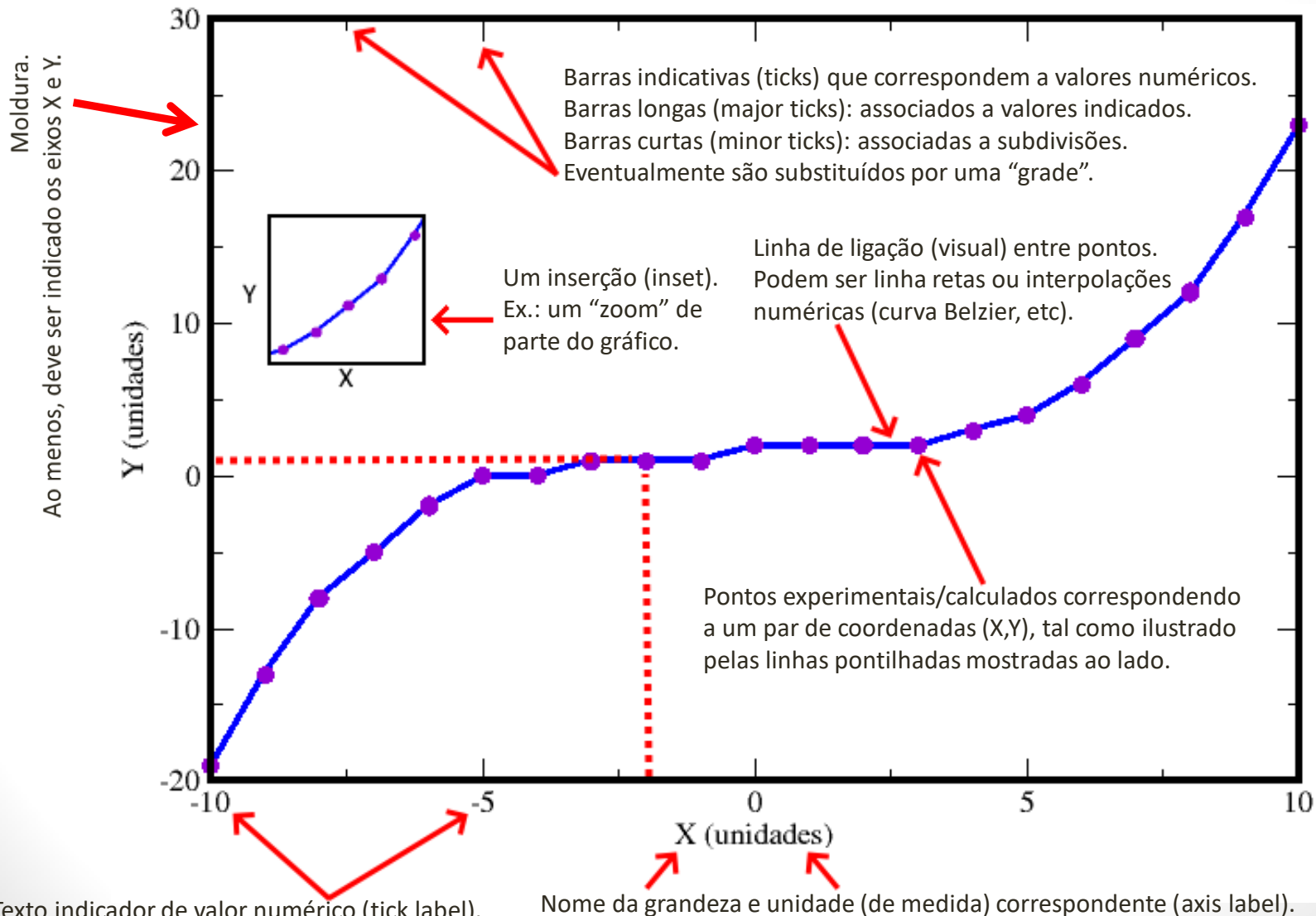
	A	B
1	X	Y
2	-10	-19
3	-9	-13,309
4	-8	-8,752
5	-7	-5,203
6	-6	-2,536
7	-5	-0,625
8	-4	0,656
9	-3	1,433
10	-2	1,832
11	-1	1,979
12	0	2
13	1	2,021
14	2	2,168
15	3	2,567
16	4	3,344
17	5	4,625
18	6	6,536
19	7	9,203
20	8	12,752
21	9	17,309
22	10	23



(*) Elaborar um gráfico é um exercício criativo que objetiva ressaltar pontos no comportamento de uma grandeza exposta como função de outra, mantendo-se fiel aos fatos, e sem distorcer a exposição dos dados/cálculos. Contudo, a interpretação de um gráfico e seu significado pode, eventualmente/geralmente, ser alvo de discussões.

Elementos básicos

É mostrado abaixo uma figura resumindo os elementos básicos de um gráfico comum, de uma grandeza, Y, como função de outra grandeza, X, baseado em dados que podem ser medidas experimentais, ou pontos calculados, ou etc



Tipos de escalas

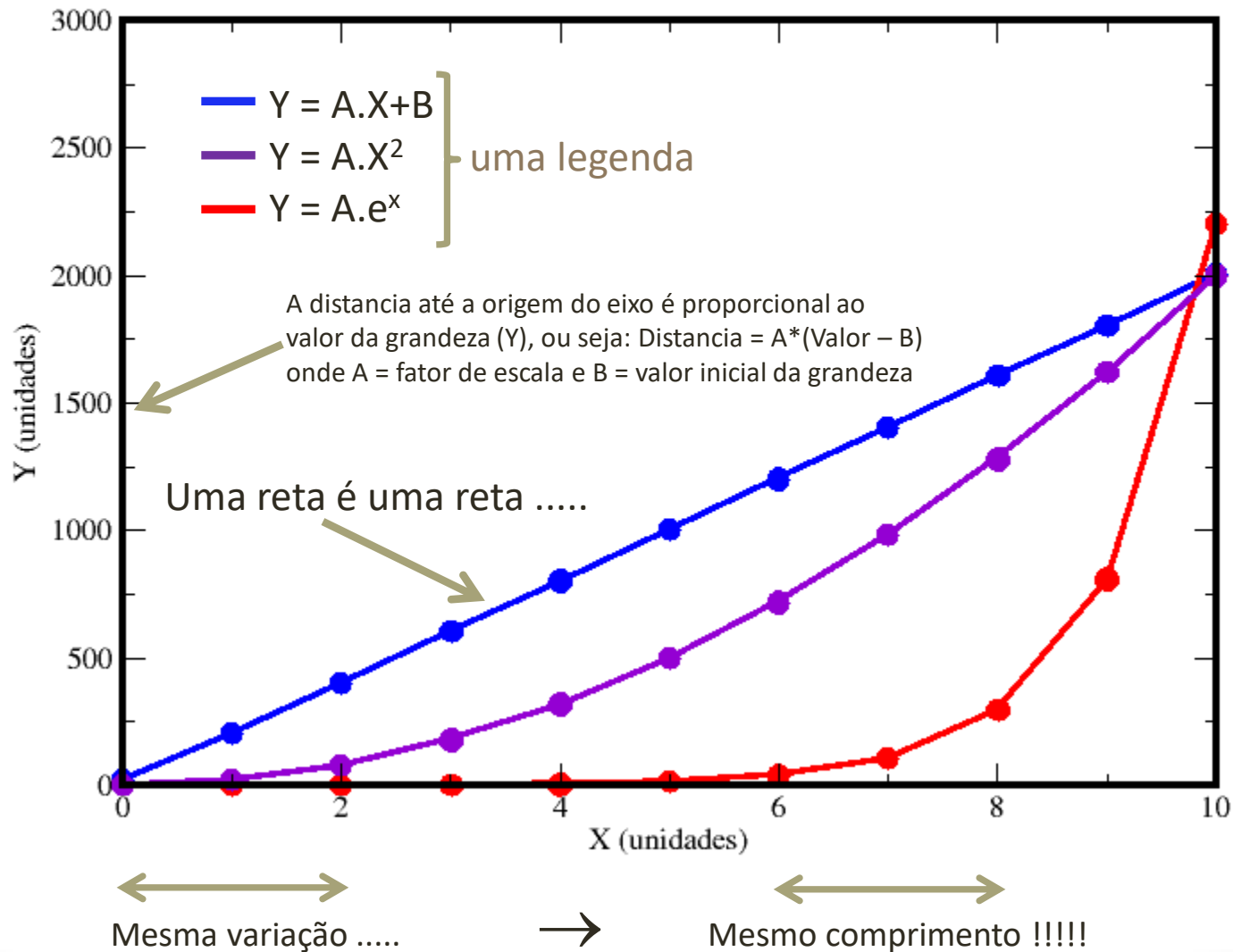
Em geral, um gráfico é um retrato de um conjunto de limitado de dados. De fato, os eixos coordenados (horizontal e vertical) são delimitados por valores máximos e mínimos. Portanto, um gráfico é usualmente um retrato de uma região de um espaço (bidimensional) definido por duas grandezas.

A escala de um gráfico pode ser definida de diversas maneiras. As mais comuns são:

- **Escala linear:** Onde o espaçamento entre valores unitários é mantida constante. Nesse caso a distancia entre a origem de um eixo até um dado ponto nesse mesmo eixo é proporcional ao valor da grandeza nesse ponto. Adicionalmente, temos consequencia que uma dada variação na grandeza considerada se mantem constante em toda a extensão do eixo.
- **Escala logarítmica:** Nesse caso o espaçamento entre valores unitários não é mantida constante. Aqui a distancia entre a origem de um eixo até um dado ponto nesse mesmo eixo é proporcional ao logaritmo do valor da grandeza nesse ponto. Em ultima analise, ao invés de “graficarmos” (plotar) a grandeza em si, “graficamos” o logaritmo dessa grandeza – e.g. não plotamos $Y \times X$ mas $\log(Y) \times X$.
- **Escala reciproca:** Aqui a distancia entre a origem de um eixo até um dado ponto nesse mesmo eixo é proporcional ao inverso do valor da grandeza nesse ponto. Em ultima analise, ao invés de “graficarmos” (plotar) uma grandeza, “graficamos” o reciproco (inverso) dessa grandeza – e.g. não plotamos $Y \times X$ mas $1/Y \times X$.

Gráfico linear/linear

É indicado abaixo os elementos básicos de um gráfico com ambas as escalas lineares, ou seja, ambas as escalas correspondentes aos eixos coordenados lineares.



O chamado “fator de escala” é uma constante, atribuída a cada um dos eixos coordenadas, que fixa a distancia no papel (medido em mm ou cm) ou no monitor de vídeo (medido em pixels), por unidade da grandeza correspondente. Tal fator de escala pode ser facilmente calculado via a expressão:

$$A = D / (\text{Max} - \text{Min})$$

onde:

A = Fator de Escala dados em (unidade de distancia)/(unidade da grandeza),

D = Distancia no papel (em mm ou cm) ou no monitor de vídeo (em pixels),

Max = Valor máximo da grandeza correspondente ao eixo em questão, e

Min = Valor mínimo da grandeza correspondente ao eixo em questão.

Por exemplo: Para um gráfico em papel com D_x milímetros para o eixo X.

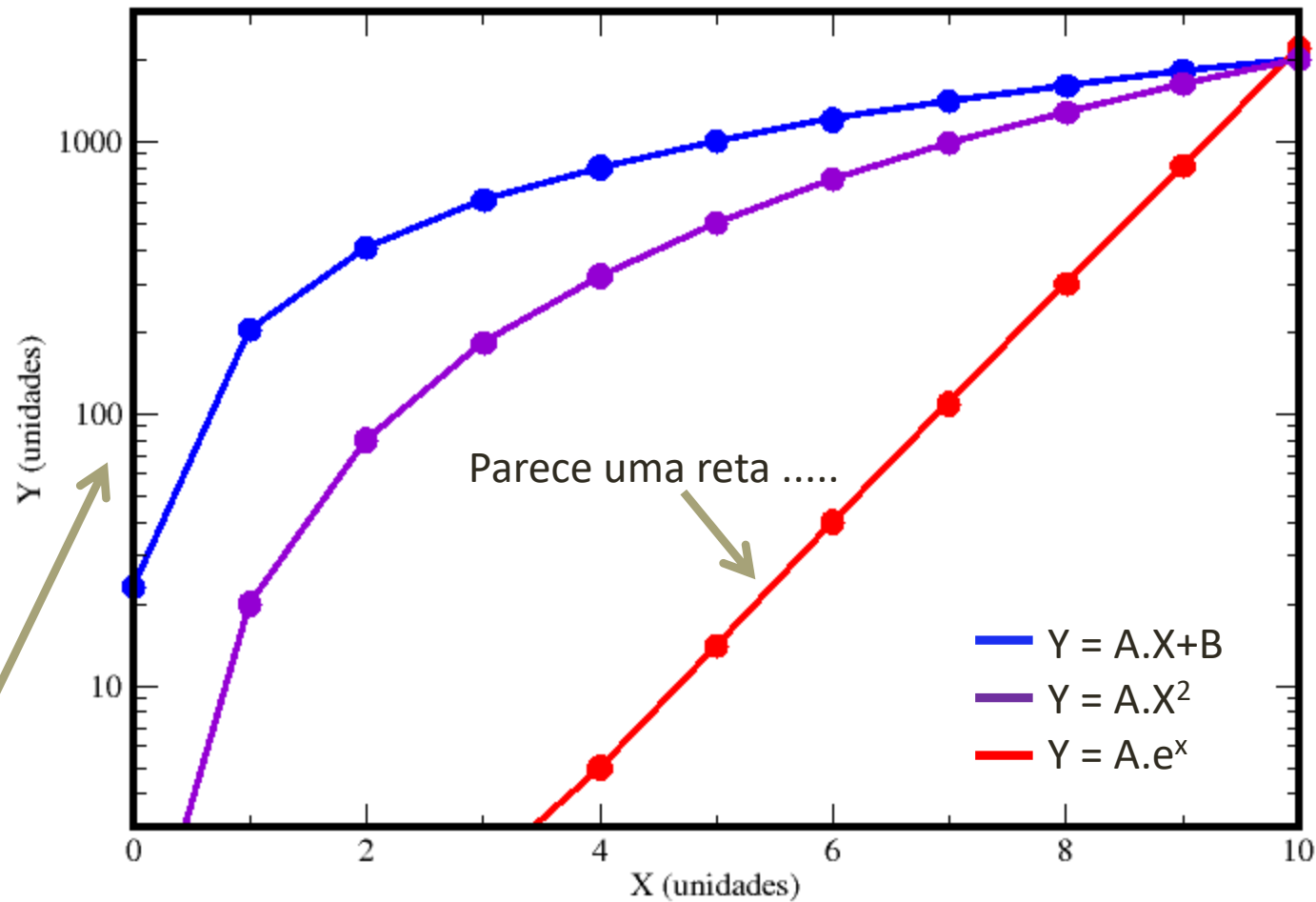
$$A_x = D_x / (X_{\text{max}} - X_{\text{min}})$$

Por exemplo: Para um gráfico em papel com D_y milímetros para o eixo Y.

$$A_y = D_y / (Y_{\text{max}} - Y_{\text{min}})$$

Gráfico linear/logarítmico

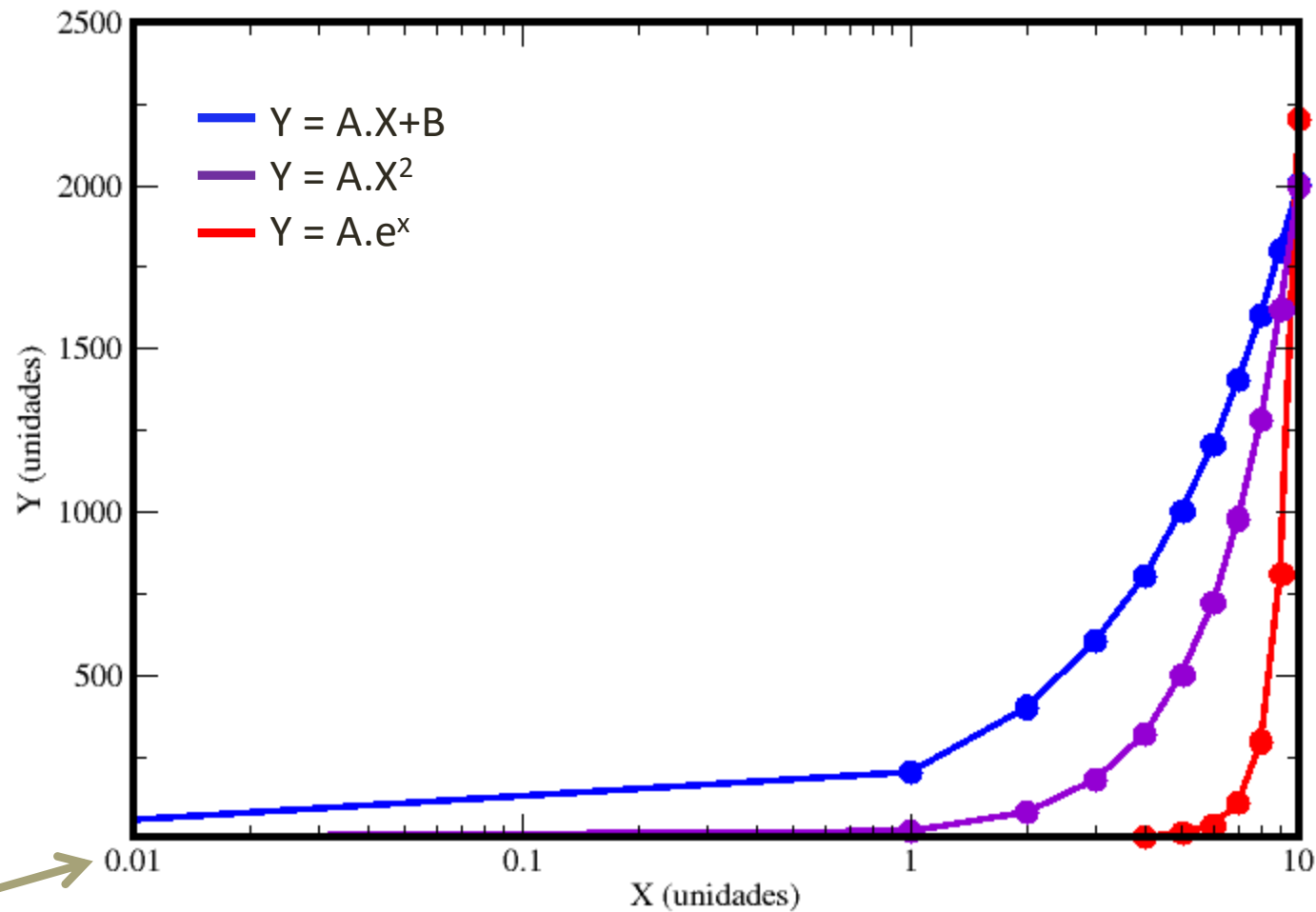
É indicado abaixo os elementos básicos de um gráfico com a escala do eixo horizontal linear e a vertical logarítmica.



A distância até a origem do eixo é proporcional ao logaritmo do valor da grandeza (Y), ou seja, estamos "graficando" $\log(Y)$ versus X. Por exemplo, a variação correspondente aos "ticks" é: 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, etc

Gráfico logarítmico/linear

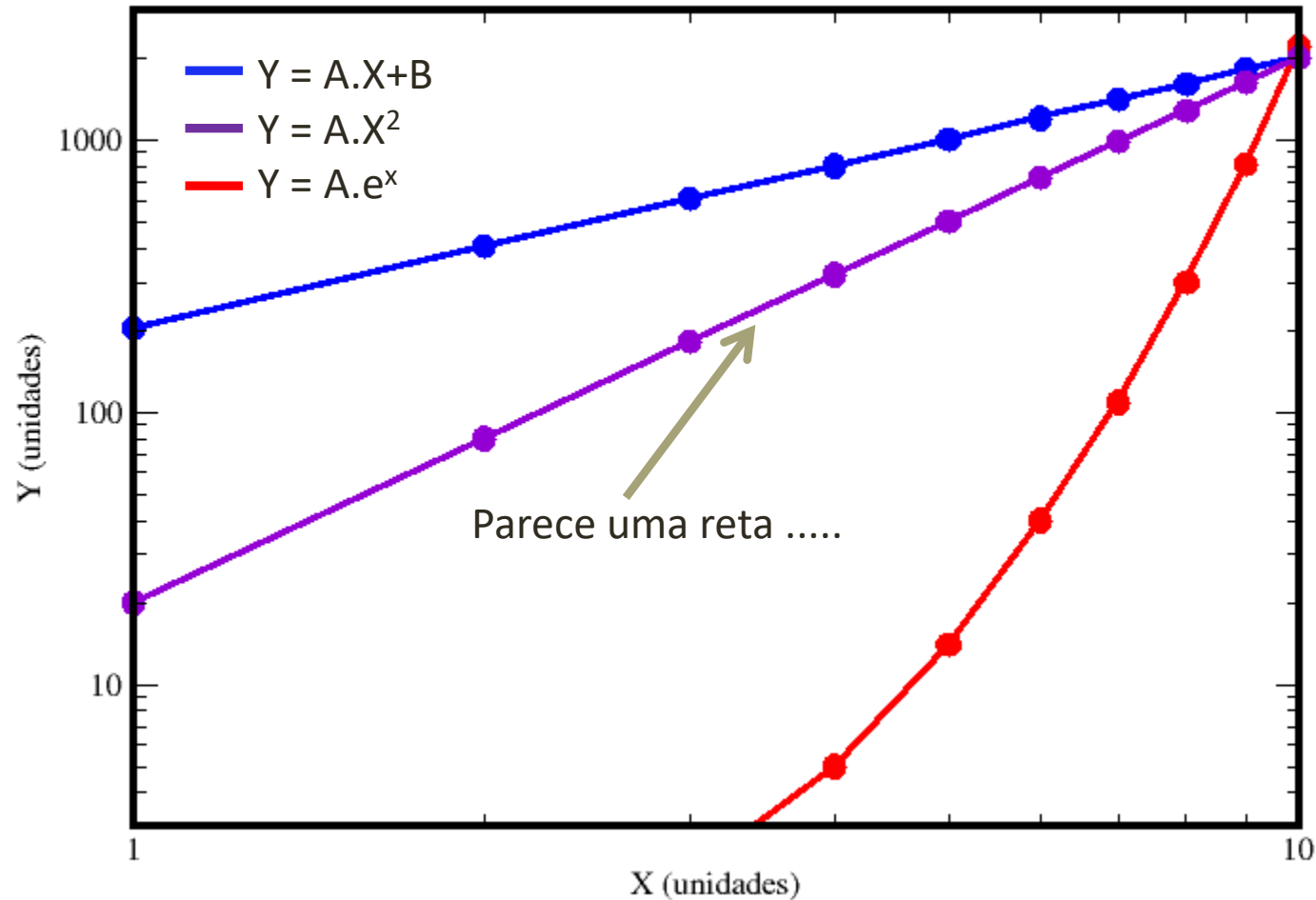
É indicado abaixo os elementos básicos de um gráfico com a escala do eixo horizontal logarítmica e a vertical linear.



A distância até a origem do eixo é proporcional ao logaritmo do valor da grandeza (X), ou seja, estamos “graficando” Y versus $\log(X)$. Por exemplo, a variação correspondente aos “ticks” é: 0.01, 0.02, 0.03,, 0.09, 0.1, 0.2, 0.3,, 1, 2, 3,, 10

Gráfico logarítmico/logarítmico

É indicado abaixo os elementos básicos de um gráfico com ambas as escalas logarítmicas, ou seja, ambas as escalas correspondentes aos eixos coordenados do tipo escala logarítmica.



A ideia por traz de gráficos lin/log e log/log

Na realidade a ideia é muito simples

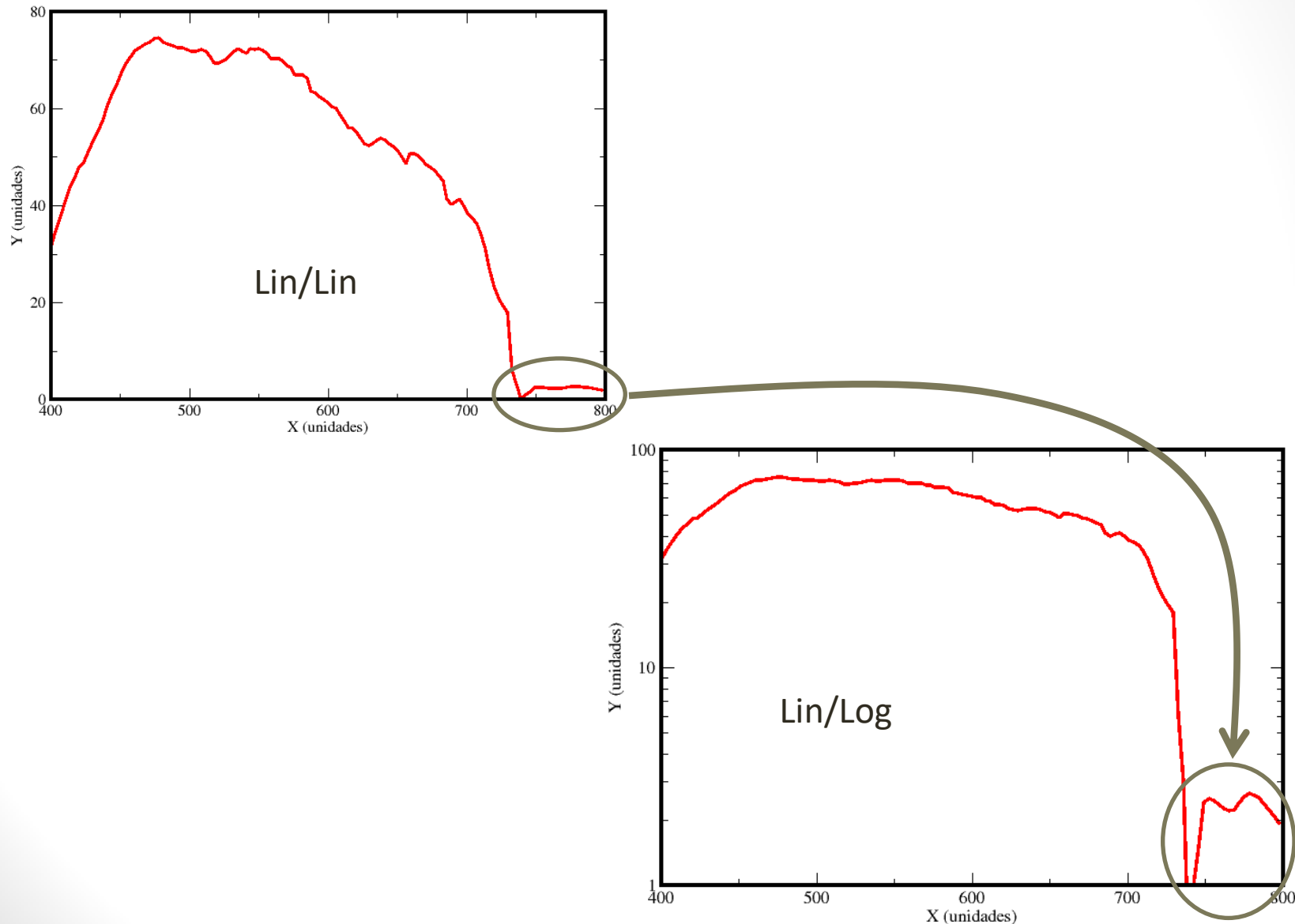
$$y = B \cdot e^{A \cdot x} \Rightarrow \text{aplicando } \ln \text{ em ambos os termos} \Rightarrow \ln(Y) = \ln(B) + A \cdot X$$

$$y = B \cdot X^A \Rightarrow \text{aplicando } \ln \text{ em ambos os termos} \Rightarrow \ln(Y) = \ln(B) + A \cdot \ln(X)$$

Varias calculadoras possuem recursos variados de regressão linear e estatística. Consulte o manual de operação de sua calculadora.

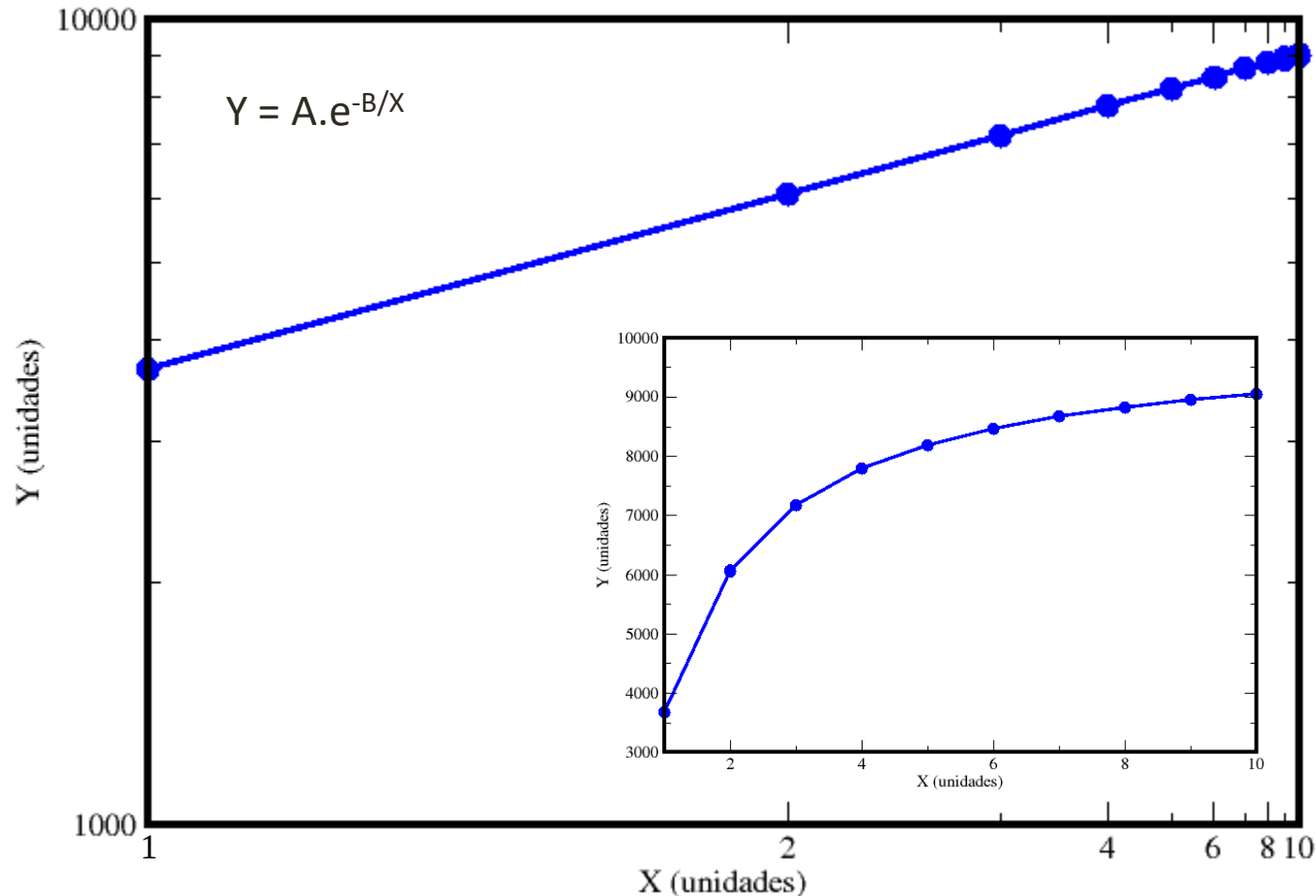
Uma outra utilidade da escala logarítmica

Esse tipo de escala é também indicada para ilustrar grandes variações de um grandeza, evidenciando “alguns” pequenos detalhes. Veja o exemplo abaixo.



Um exemplo usando escala reciproca

É indicado abaixo um exemplo desse tipo de escala em um gráfico tipo Arrhenius.



Aqui é “graficado” $\log(Y)$ versus $1/X$. Contudo, o eixo horizontal é um reflexo em espelho !!!!! Isso ilustra o quanto podemos ser engenhosos para fazer uma dependência funcional mais complexa parecer uma reta

Softwares para confecção de gráficos

É indicado abaixo alguns softwares usados na elaboração de gráficos diversos.

Aplicativos proprietários (necessita licença paga)

- Microsoft Excel (planilha de cálculo que é parte do Microsoft Office)
- Microcal Origin (somente Windows, somente com “assinatura anual”)
- Etc, etc, e tal

Aplicativos de código aberto (livres e gratuitos)

- LibreOffice (em particular a planilha Calc, disponível no Linux, Windows, etc)
- **GNU Plot** (usa linha de comando, disponível no Linux, Windows, etc)
- SCI Davis (disponível no Linux e Windows)
- LabPlot (disponível no Linux e Windows)
- Xm-Grace (disponível somente no Linux, mas não inclui planilha de dados
- **Qt-Grace** (disponível para Windows, mas não inclui planilha de dados
- Etc, etc, e tal

Regressão linear do tipo $Y = A.X + B$

Consideremos um conjunto de dados composto por N pares coordenados (X_i, Y_i) , com $i = 1, 2, 3, \dots, N$. Se desejamos ajustar uma reta a esse grupo de dados, podemos calcular os coeficientes A (inclinação da reta) e B (termo independente), via as seguintes expressões:

$$Y = A.X + B$$

$$A = \frac{N \cdot (\sum_{i=1}^N X_i \cdot Y_i) - (\sum_{i=1}^N X_i) \cdot (\sum_{i=1}^N Y_i)}{N \cdot (\sum_{i=1}^N X_i^2) - (\sum_{i=1}^N X_i)^2}$$

$$B = \frac{(\sum_{i=1}^N Y_i) \cdot (\sum_{i=1}^N X_i^2) - (\sum_{i=1}^N X_i) \cdot (\sum_{i=1}^N X_i \cdot Y_i)}{N \cdot (\sum_{i=1}^N X_i^2) - (\sum_{i=1}^N X_i)^2}$$

Regressão linear do tipo $Y = A.X$

Consideremos um conjunto de dados composto por N pares coordenados (X_i, Y_i) , com $i = 1, 2, 3, \dots, N$. Se desejamos ajustar uma reta a esse grupo de dados, podemos calcular o coeficiente A (inclinação da reta), via a expressão:

$$Y = A.X$$

$$A = \frac{(\sum_{i=1}^N X_i \cdot Y_i)}{(\sum_{i=1}^N X_i^2)}$$

Cálculo do fator de correlação

Consideremos um conjunto de dados composto por N pares coordenados (X_i, Y_i) , com $i = 1, 2, 3, \dots, N$. Após ajustar uma reta a esse grupo de dados, podemos calcular um segundo grupo de valores $Y'_i = A.X_i + B$, com base no mencionado ajuste. Dados esses dois conjuntos de dados podemos calcular o chamado coeficiente de correlação, r , que fornece uma “estimativa” de o quão confiável é o referido ajuste. Em outras palavras, o valor de $R_{\%} = 100.r$, pode ser traduzido como a probabilidade percentual do grupo de dados original, (X_i, Y_i) , ser representado pela reta obtida via regressão linear.

$$r = \text{sgn}(a) \cdot \sqrt{\frac{\sum_{i=1}^N (Y'_i - \bar{Y})^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2}} \quad \text{com} \quad \bar{Y} = \frac{\sum_{i=1}^N Y_i}{N}$$

onde $\text{sgn}(A)$ é a chamada função sinal definida por:

$\text{sgn}(A)=1$ para $A>0$

$\text{sgn}(A)=-1$ para $A<0$

Quanto mais o valor de r for próximo de 1 ou -1 (dependendo do sinal de A), mais fidedigno e confiável será o ajuste da reta aos dados originais. Em geral, os valores aceitáveis para um bom sucesso de uma regressão linear é $|r| > 0,95$.