Theory & Notation Guidelines

Please contact your editor and/or William Poole if you have any questions

June 19, 2021

1 Introduction

This document provides notation guidelines for commonly used mathematics in *The Art of Molecular Programming* Community Textbook. The goal of this document is not to define all the quantities, symbols, and notation used in the textbook but instead to provide rough guidelines for the kinds of symbols to use in different situations.

2 Variables and Functions

- **Scalars, Vectors, and Matrices**: there will be no notational difference between scalars, vectors, and matrices (e.g. $\vec{x}$ will not be used). Instead, if variables are meant to be vectors or matrices this should be written in the text and made explicit in formulas via the use of one or more indices.
- **Functions:** when used generally should be denoted f or g. Specific functions may use different symbols. Use parentheses not brackets, eg $f(x, y)$ instead of $f[x, y]$.
- **Indices & subscripts:** Use i and j followed by l, m, n. *Please avoid k as an index.*
- **Superscripts:** superscripts will be assumed to mean exponentiation.
- **State Variables:** molecular counts, concentrations, states of a computational model, and other dynamic quantities should be lower-case s, x, y, and/or z. State variables may be vectors $x = (x_1...x_i...x_n)$ with subscripts used to index the components. Time dependence should be shown via parenthesis e.g. $x(t)$.
- **Chemical Species:** the formal representations of molecules should be capitalized versions of their dynamic quantities. For example, the species S_i is at concentration s_i.
- **Rates and Constants:** k and K for example K_D (Dissociation Constant), K_A (Association Constant), k_b (binding rate), unbinding rate k_u (unbinding rate), k_1 (rate 1), k_i (rate i), etc. Other constants in formulas can be drawn from **additional symbols**. Rates and constants may also be vectors or matrices with subscripts used to index the components.
- **Spaces:** (e.g. state spaces and reachability classes). Use Γ, Γ', and additional symbols as needed. *Please avoid abstract mathematical spaces such as Vector Spaces or Hilbert Spaces.*
- **Sets of Numbers:** use $\mathbb{R}$ and $\mathbb{Z}$ to denote the real numbers and integers, respectively, and $\mathbb{R}_{\geq 0}$ or $\mathbb{Z}_{\geq 0}$ to denote the non-negative numbers. *Complex or other abstract number systems should not be necessary for this text.*
- **Additional Symbols:** The letters a, b, c, d, A, B, C, D, and the Greek letters $\alpha, \beta, \gamma, \delta$ can also be used for constants, functions, or variables as needed.

3 Computer Science

- **Algorithms and Pseudocode:** all algorithms should be written in psuedocode (avoiding language specific jargon and notation) and include descriptions chosen for clarity over brevity (in other words, explain what the code is doing in words). Pseudocode should be monospaced using \textttt{ ... } and indented from the main text.
- **Graphs:** use the notation $G = (V, E)$ where G is a graph with vertices V and edges E.
- **Logic:** use: Please write out logic functions using monospaced font \textttt{ ... } e.g. `NOT`, `AND`, `OR`, `XOR`, etc..
- **Complexity:** use big-O notation whenever possible: e.g. $\mathcal{O}(n)$ denotes linear scaling. For lower bounds, please use Ω notation. When including any bound, please state explicitly in words if it is upper, lower, scaling, etc. *In general, please try to avoid other kinds of bounds such as little-o, etc. If these are absolutely essential to a section, please discuss how to use and define these bounds in the text with your editor.*

4 Sequences, Structures, and Strands

Please write sequences and strands using monospaced font. In latex, this can be written using \texttt{...}.

- **DNA Sequences:** explicitly denote 5′ and 3′ end and use uppercase monospaced font e.g. `5′-ACTG-3′`.
- **RNA Sequences:** explicitly denote 5′ and 3′ ends and use lowercase monospaced font e.g. `5′-acug-3′` (Note the use of the
prime symbol instead of an apostrophe).
- **Wildcard Bases:** Use `N` (DNA) and `n` (RNA) for wildcard bases. If more specific wild-card base types are needed, please use the IUB standard (link).
- **Peptide Sequences:** use standard three-letter code and show the N/C termini in a monospaced font e.g. N-Argenine-Hystidine-Lysine-C would be written. `N-arg-hys-lys-C`.
- **Domain Level Representation:** domains can use any unrestricted lowercase variable such as $a, b, c, x, y...$ to represent a sequence of arbitrary length. In some constructions, it may be convenient to use an index e.g. x_i to represent different domains.
- **Complementary Domains:** the complement to the domain x should be written as x^*.
- **Secondary Structure:** we recommend showing secondary structure graphically. However, if a notation is needed we suggest use the dot-parens-plus (DU+) standard in a mono-spaced font. An example can be found here.
- **Genetic Circuit Representations:** Use (SBOL 3) glyphs to represent genetic circuit if possible.
- Genes and Proteins: The name of a gene should be italicized and the protein should not, for example *LacZ* produces β-galactosidase.

5 Calculus

- **Discrete Changes:** use Δx to denote the change in a variable x.
- **Derivatives:** use $\frac{\mathrm{d}x}{\mathrm{d}y}$ and to denote the (non-partial) derivative of x with respect to y. In latex, $\backslash dv\{x\}\{y\}$. *Please avoid other notations such as $\dot{x}$ or $\partial_t X$*
- **Partial Derivatives:** use $\frac{\partial x}{\partial y}$ to denote the partial derivative of x with respect to y. In latex, $\backslash pdv\{x\}\{y\}$. When considering a partial derivative where a third variable z is held constant, please use the notation $\frac{\partial x}{\partial y}\big|_z$.

- **Integrals:** use $\int_0^x g(x')\,\mathrm{d}x'$ to denote the integrals of $g(x')$ from 0 to x. *Please avoid advanced vector calculus integrals such as surface or path integrals.*
- **Gradient and Laplacian**: use ∇ (and explicitly write out the derivatives) to denote the gradient and ∇^2 (and explicitly write out the derivatives) to denote the Laplacian.

6 Probability and Statistics

- **Probability:** denote the probability of an event x with $\mathbb{P}(x)$. To denote conditional probability please use $\mathbb{P}(x \mid y)$.
- **Expected Value:** denote the expected value of a variable x with respect to a distribution P as $\mathbb{E}_P(x)$. *Please avoid the notation $\langle ... \rangle$ for expected value.*
- **Variance and Standard Deviation:** please use σ^2 to denote variance and σ to denote standard deviation.
- **Distributions:** use P and Q to denote general probability distributions. For named distributions it is okay to use symbols that correspond their name; for example, $\mathcal{N}$, $\mathcal{P}$, and $\mathcal{B}$, for normal distributions, Poisson distributions, and binomial distributions, respectively.
- **Uncertainty in a single experimental measurements:** error bars and experimental variance indications in text will be assumed to denote standard error $SE = \sigma/\sqrt{N}$ unless it is explicitly noted otherwise. For example, given data with a mean a and standard error b and sample size n write $a \pm b\,(N = n)$. *When quoting results or re-purposing a figure with a different convention without the necessary data to calculate standard error, explicitly state the error used.*
- **Uncertainty between multiple experiments:** in this case, we are referring to reporting measurements across multiple experiments (technical replicates or different methodologies to measure the same quantity) please report standard deviation the mean: $SDM = \frac{\sigma}{\sqrt{N}}$ where σ is the standard deviation between the N total experiments.

7 Physics & Thermodynamics

- **Physical Quantities:** ρ (density), η (viscosity), L (length), N (count), V (volume), T (temperature), m (mass)
- **Thermodynamic Functions:** Due to the numerous thermodynamic variables that might be necessary and to ensure they do not inhibit authors from using these letters in other sections, thermodynamic quantities will be written in script (latex \mathcal{ ... }); $\mathcal{G}$ (free energy), E (energy), $\mathcal{U}$ (internal energy), $\mathcal{H}$ (enthalpy), $\mathcal{S}$ (entropy), $\mathcal{Q}$ (heat), $\mathcal{Z}$ (partition function), μ (chemical potential)
- **Constants:** k_B (Boltzmann Constant), R (ideal gas constant)
- **Intensive vs Extensive:** If a quantity (capital letter) is extensive, the symbol used for the specific quantity (divided by mass) should be the lower case. (e.g. specific heat capacity, $c_p = C_p/m$)
- **Other quantities and functions:** follow the IUPAC standard (link) (cheat-sheet link) whenever these recommendations don't clash with the rest of this guide.

8 Chemistry

- **Elements and Chemical Names:** please ensure elements and the names of chemicals are *not italicized*, eg H_2 instead of H_2. We reocmmend the mhchem package in Latex for chemical species and reactions.