Notation for CS395T: Continuous Algorithms

Kevin Tian

General notation. We use $[d]$ to denote the set $\{i \in \mathbb{N} \mid i \le d\}$. We let $\iota := \sqrt{-1}$ denote the imaginary unit. We use $s \sim_{\text{unif.}} S$ to denote a uniform sample from the set S. When S is a subset of T clear from context, we let $S^c := T \setminus S$ denote its complement. We denote vectors in lowercase boldface letters. When $\mathbf{v}$ is a vector, we refer to its i^{th} coordinate by $\mathbf{v}_i$, and if the vector has a subscript, e.g., it is a variable $\mathbf{v}_t$, we denote its i^{th} coordinate by $[\mathbf{v}_t]_i$. We use $\approx$, $\gtrsim$, and $\lesssim$ to hide universal constants, e.g. $x \lesssim y$ means there is a universal constant C such that $x \le Cy$. We use $\mathbf{1}_d$ and $\mathbf{0}_d$ to denote the all-ones and all-zeroes vectors of dimension d respectively. We use $\widetilde{O}$ to hide polylogarithmic factors in problem parameters for simplicity.[1] We let $\text{supp}(\mathbf{v})$ denote the support of a vector $\mathbf{v} \in \mathbb{R}^d$, i.e., the subset of coordinates $i \in [d]$ where $\mathbf{v}_i \neq 0$. For $x \in \mathbb{R}$, we let $\text{sign}(x) := 1$ if $x \ge 0$, and otherwise we let $\text{sign}(x) := -1$. We let the i^{th} standard basis vector in $\mathbb{R}^d$ be denoted by $\mathbf{e}_i$, i.e., the 0-1 vector with $\text{supp}(\mathbf{e}_i) = \{i\}$.

Matrices. We denote matrices in uppercase boldface letters. We let $\mathbf{I}_d$ denote the $d \times d$ identity matrix, and $\mathbf{0}_{m\times n}$ be the $m \times n$ all-zeroes matrix. We let $\mathbb{S}^{d\times d}$ be the set of symmetric $d \times d$ matrices, which we equip with $\preceq$, the Loewner partial ordering (i.e., $\mathbf{M} \preceq \mathbf{N}$ implies $\mathbf{N} - \mathbf{M}$ is positive semidefinite). We also let $\mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ denote the subset of $d \times d$ positive semidefinite matrices, and $\mathbb{S}^{d\times d}_{\succ \mathbf{0}}$ are the $d \times d$ positive definite matrices. The number of nonzero entries of a matrix $\mathbf{M}$ is denoted $\text{nnz}(\mathbf{M})$. We let $\mathcal{T}_{\text{mv}}(\mathbf{M})$ be the time it takes to compute $\mathbf{M}\mathbf{v}$ for an arbitrary vector $\mathbf{v}$;[2] note that $\mathcal{T}_{\text{mv}}(\mathbf{M}) = O(\text{nnz}(\mathbf{M}))$, and if $\mathbf{M} \in \mathbb{R}^{m\times n}$ is given by a rank-k decomposition $\mathbf{M} = \mathbf{U}\mathbf{V}^\top$, we have $\mathcal{T}_{\text{mv}}(\mathbf{M}) = O((m+n)k)$. We let $\omega \approx 2.372$ be the current matrix multiplication exponent, i.e., such that we can multiply two $d \times d$ matrices in $O(d^\omega)$ time. When $\mathbf{M} \in \mathbb{S}^{d\times d}$ has eigendecomposition $\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ and f is a real-valued function whose domain contains the spectrum of $\mathbf{M}$, we overload $f(\mathbf{M}) := \mathbf{U}f(\mathbf{\Lambda})\mathbf{U}^\top$ where $f(\mathbf{\Lambda})$ is applied entrywise on the diagonal. We reserve $\|\cdot\|_{\text{op}}$, $\|\cdot\|_{\text{tr}}$, and $\|\cdot\|_{\text{F}}$ for the operator norm, trace norm, and Frobenius norm of a matrix (a.k.a. the ∞-, 1-, and 2-Schatten norms). When $\mathbf{T}$ is a k-way tensor operating on vector inputs $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$, we write $\mathbf{T}[\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$ to mean the resulting scalar from this operation. When we drop some set of $\ell \in [k]$ of the inputs (with ordering clear from context), we mean the ℓ-way tensor operating on the remaining inputs, e.g., $\mathbf{T}[\mathbf{v}_1]$ is a $(k-1)$-way tensor. For example, $\mathbf{M}[\mathbf{u}, \mathbf{v}] = \mathbf{u}^\top \mathbf{M}\mathbf{v}$ when $\mathbf{M}$ is a matrix, and $\mathbf{M}[\mathbf{u}] = \mathbf{M}^\top \mathbf{u}$. We let $\text{Span}(\mathbf{M})$ denote the span of the columns of $\mathbf{M}$, and $\text{rank}(\mathbf{M})$ denote its rank.

Norms. We let $\|\cdot\|$ denote a norm on $\mathbb{R}^d$. For a norm $\|\cdot\|$ on $\mathbb{R}^d$, we let $\|\cdot\|_*$ denote the dual norm. When applied to a vector or matrix argument, $\|\cdot\|_p$ denotes the ℓ_p or Schatten-p norm respectively. For $\mathbf{v} \in \mathbb{R}^d$ and $r > 0$, if $\|\cdot\|$ is a norm on $\mathbb{R}^d$, we let $\mathbb{B}_{\|\cdot\|}(\mathbf{v}, r) := \{\mathbf{v}' \in \mathbb{R}^d \mid \|\mathbf{v}' - \mathbf{v}\| \le r\}$ denote the associated norm ball around $\mathbf{v}$. When $\|\cdot\|$ is omitted, we always assume $\|\cdot\| = \|\cdot\|_2$, and when $\mathbf{v}$ is omitted, we always assume $\mathbf{v} = \mathbf{0}_d$. For a matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$ and $p, q \ge 1$, we define

$$\|\mathbf{M}\|_{p\to q} := \max_{\|\mathbf{v}\|_p \le 1} \|\mathbf{M}\mathbf{v}\|_q .$$

Sets. We let χ_S be the 0-∞ indicator of a set S, such that

$$\chi_S(x) = \begin{cases} 0 & x \in S \\ \infty & x \notin S \end{cases}.$$

[1] This usage of $\widetilde{O}$ (without declaring what polylogarithmic factors are hidden) is somewhat controversial in the community, but it significantly saves on space for some very hairy theorem statements. I promise I will declare if anything particularly nefarious is being hidden by $\widetilde{O}$; otherwise, it should be reasonable from context clues.

[2] If $\mathbf{M} \in \mathbb{R}^{n\times d}$, we usually assume for simplicity that $\mathcal{T}_{\text{mv}}(\mathbf{M}) = \Omega(n + d)$, as we must at least process the input and write down the output. If $\mathbf{M}$ has all-zero columns or rows, we can first drop them and reduce the dimension.

For a set $S \subseteq \mathbb{R}^d$ and $\lambda \in \mathbb{R}$, we write $\lambda S := \{\lambda \mathbf{v} \mid \mathbf{v} \in S\}$, $S^c := \{\mathbf{v} \in \mathbb{R}^d \mid \mathbf{v} \notin S\}$, and $\mathrm{Vol}(S)$ denotes the volume (Lebesgue measure) of S in $\mathbb{R}^d$. We denote the Minkowski sum of sets by $\oplus$, i.e., $A \oplus B := \{\mathbf{v} \mid \mathbf{v} = \mathbf{a} + \mathbf{b}, \mathbf{a} \in A, \mathbf{b} \in B\}$. We use $\mathrm{Conv}(S)$ to mean the convex hull of a set S, and $\mathrm{relint}(S)$ to mean the relative interior of S. For $S \subseteq \mathbb{R}^d$, we let $\mathbf{\Pi}_S(\mathbf{v}) := \mathrm{argmin}_{\mathbf{v}' \in S} \|\mathbf{v} - \mathbf{v}'\|_2$ denote the Euclidean projection of $\mathbf{v}$ to S.

Functions. When f is a function on some decision variable $\mathbf{x}$, we sometimes use $\cdot$ in place of the argument $\mathbf{x}$ to denote the function itself, e.g., $\|\cdot\|$ denotes the function which, when evaluated at $\mathbf{x}$, returns $\|\mathbf{x}\|$. When integrating a function f without specifying a domain of integration, we always mean the entire domain of f. We use ∇^k to denote the k^{th} derivative tensor of a k-times differentiable multivariate function, e.g., ∇f is the gradient of differentiable $f : \mathbb{R}^d \to \mathbb{R}$. In one dimension this is denoted $f^{(k)}$.

Probability. Expectations of random variables, denoted $\mathbb{E}$, are always taken with respect to all randomness used to define the variable unless otherwise specified. For a scalar random variable Z we let $\mathrm{Var}[Z] := \mathbb{E}[Z^2] - (\mathbb{E}Z)^2$ denote its variance. When $\mathcal{E}$ is an event on a probability space clear from context, we let $\mathbb{I}_{\mathcal{E}}$ denote the random 0-1 variable which is 1 iff $\mathcal{E}$ occurs. When μ is a probability density, we write $x \sim \mu$ to denote a sample from this density. We denote the support of a distribution $\mathcal{D}$, i.e., all values samples from $\mathcal{D}$ can take on, by $\mathrm{supp}(\mathcal{D})$. When f is a nonnegative integrable function, we write $\mu \propto f$ to mean the density taking on values $\frac{f}{Z}$, where $Z = \int f(x)\mathrm{d}x$ is the normalizing constant. We let $\mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ denote the multivariate Gaussian distribution with specified mean $\boldsymbol{\mu} \in \mathbb{R}^d$ and covariance $\mathbf{\Sigma} \in \mathbb{S}^{d \times d}_{\succeq \mathbf{0}}$. For two distributions P, Q, we let $\Gamma(P, Q)$ denote the set of couplings of P and Q.