Towards a Compositional Framework for Convex Analysis

Dario Stein
Radboud Universiteit Nijmegen
dario.stein@ru.nl

Richard Samuelson
Humming, Inc.
richard@heyhumming.com

We outline a categorical framework for convex analysis based on the notion of *convex bifunction*, highlighting connections with categorical probability.

Convex analysis is a classical area of mathematics with numerous applications in optimization, economics, physics, statistics and information theory. The central notion is that of a convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ into the extended reals $\overline{\mathbb{R}} = [-\infty, +\infty]$, meaning its epigraph $\mathrm{epi}(f) = \{(x,t) : t \geq f(x)\}$ is a convex subset of $\mathbb{R}^{n+1}$. Convex analysis admits a beautiful duality theory; the ubiquitous Legendre-Fenchel transform (or convex conjugation), defined as

$$f^*(x^*) = \sup_x \{\langle x^*, x\rangle - f(x)\}$$

decomposes f in terms of all affine function $\langle x^*, x\rangle - c$ majorized by f. The function f^* is itself convex, and under a closedness assumption we recover $f^{**} = f$. A useful way to combine convex functions is *infimal convolution*

$$(f \mathbin{\Box} g)(x) = \inf_y \{f(x-y) + g(y)\}$$

This operation is dual to addition, in the sense that $\mathrm{epi}(f \mathbin{\Box} g) = \mathrm{epi}(f) + \mathrm{epi}(g)$, and convex conjugation interchanges the two $(f \mathbin{\Box} g)^* = f^* + g^*$, $(\mathrm{cl}(f) + \mathrm{cl}(g))^* = \mathrm{cl}(f^* \mathbin{\Box} g^*)$ where $\mathrm{cl}(f) = f^{**}$ is lower-semicontinuous closure.

Convex functions are not closed under composition. In order phrase convex analysis in compositional terms, we invoke the notion of *convex bifunction* $F : \mathbb{R}^m \rightharpoonup \mathbb{R}^n$, which is a curried convex function $F : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ [12]. Convex bifunctions compose as follows

$$(F \circ G)(x,z) = \inf_y \{F(y,z) + G(x,y)\}$$

and convex functions $\mathbb{R}^m \to \overline{\mathbb{R}}$ are recovered as effects $\mathbb{R}^m \rightharpoonup I$. The category CxBiFn has the structure of a hypergraph category [4], following the general construction [1, 9, 3] of relations valued in the commutative quantale $\overline{\mathbb{R}}$ [7, 15]. Convex conjugation can be extended to bifunctions and under semicontinuity assumptions defines a contravariant equivalence with $(F \circ G)^* = G^* \circ F^*$ [12, § 38].

The goal of this talk is to clearly underline the compositional structure of convex analysis. The field is full of exceptions and caveats (like the notion of properness) which makes a high level description tricky. We propose to overcome this in two ways: (a) by extending the enriched-categorical treatment of [15] to convex bifunctions, and (b) by identifying well-behaved subcategories of bifunctions in which composition and duality work as desired. We elaborate some examples of interest:

Linear Algebra Let Vect denote the spaces $\mathbb{R}^n$ and linear maps $A : \mathbb{R}^m \to \mathbb{R}^n$. We have a faithful hypergraph embedding $F : \mathrm{Vect} \to \mathrm{CxBiFn}$ via indicator bifunctions[1] $F_A(x,y) = \{\!|y = Ax|\!\}$. Not only do these bifunctions compose like linear maps, convex duality generalizes the dualities of linear algebra [2]: the dual F_A^* is the the indicator bifunction of the adjoint map A^*. [12, § 30].

[1] we use the variant of Iverson brackets appropriate for convex analysis, i.e. $\{\!|\varphi|\!\} = 0$ if φ, otherwise $+\infty$

Submitted to:
ACT 2023

Convex Relations A convex relation $R : \mathbb{R}^m \nrightarrow \mathbb{R}^n$ is a convex subset of $R \subseteq \mathbb{R}^{m \times n}$ [1, 3, 9]. Convex relations compose using ordinary relation composition, and indicator bifunctions $F_R(x,y) = \{(x,y) \in R\}$ again define a hypergraph functor into CxBiFn. The duals of indicator functions are well-known as support functions [12]. Conversely, we may consider the Writer monad $T(X) = X \otimes \mathbb{R}$ on $\texttt{CxRel}$ for the monoid $(\mathbb{R}, +, 0)$. Every convex bifunction $F : \mathbb{R}^m \rightharpoonup \mathbb{R}^n$ becomes a Kleisli arrow via its epigraph $\mathrm{epi}(F) : \mathbb{R}^m \nrightarrow T(\mathbb{R}^n)$, and bifunction composition coincides with Kleisli composition in a wide range of circumstances, with the precise relationship remaining to be clarified.

Probability Theory and Laplace Approximation Many widely used probability distributions (Gaussian, Laplace, Dirichlet, Exponential, Uniform) are logconcave, meaning their logdensity $h(x) = \log f(x)$ is concave. Using the formalism of Markov categories [6], we can naturally see categories of concave bifunctions[2] as abstract probability theories and study their relationship to measure-theoretic probability categorically. Composition is maximization (mode-finding) instead of integration: This can be seen as an approximation (Laplace approximation, method of steepest descent) to ordinary probability, that is we approximate an integral

$$\int_a^b \exp(h(x))\mathrm{d}x \approx \int_a^b \exp\left(h(x_0) + \frac{1}{2}h''(x_0)(x-x_0)^2\right)\mathrm{d}x$$

by fitting a Gaussian density around its mode x_0 (e.g. [11]). The Laplace aproximation is a crucial tool for example in predictive coding [5]. Convolution of probability densities $\int f_X(x) g_Y(z-x)\mathrm{d}x$ is approximated by the inf-convolution of logdensities. More formally the tropical semiring $(\overline{\mathbb{R}}, \max, +)$ can be seen as a limit or deformation of $[0,\infty]$ under taking logarithms. This process is known as Maslov dequantization [8] and gives rise to the field of idempotent mathematics, where the Legendre transform becomes analogous to the Fourier transform in real analysis.

It is not surprising that the Laplace approximation is exact for Gaussian distributions; calculations with Gaussian densities are equivalent to convex algebra with their negative logdensity counterparts, which are quadratic functions. This defines a monoidal embedding from the Markov category $\texttt{Gauss}$ [6] into convex bifunctions (modulo scalars). The Legendre transform here takes the form of the duality between densities (precision) and covariance.

By working with convex functions, a variety of phenomena becomes easier to define: For example, there can be no measure-theoretic uniform distribution over $\mathbb{R}$; the improper Lebesgue density 1 is not normalizable, and the convolution $1 * 1 = +\infty$ diverges. The negative logdensity 0 on the other hand is normalized in CxBiFn, and the inf-convolution $0 \,\Box\, 0 = 0$ works as desired. This allows us to define a hypergraph category of Gaussian relations [14] which includes Gaussian probability as well as linear relations (uninformative priors) purely in terms of convex analysis, where is corresponds to partial quadratic convex functions (pqcfs). Conjugates of pqcfs are again pqcfs, generalizing the precision-covariance duality from earlier [12, § 12].

More generally, for a random variable X, the cumulant-generating function $c_X(s) = \log \mathbb{E}[\exp(sX)]$ is always convex, and its convex conjugate c_X^* is known as *rate function*, which has applications in large deviations theory and variational inference [10, 16]. The notion of cumulant-generating function is central to the study of exponential families, which seem like a promising direction for generalizing the current theory.

[2] modulo additive constants, see *automatic normalization* in [13]

References

[1] Joe Bolt, Bob Coecke, Fabrizio Genovese, Martha Lewis, Dan Marsden & Robin Piedeleu (2019): *Interacting conceptual spaces I: Grammatical composition of concepts*. Conceptual spaces: Elaborations and applications, pp. 151–181.

[2] Filippo Bonchi, Paweł Sobociński & Fabio Zanasi (2017): *Interacting hopf algebras*. Journal of Pure and Applied Algebra 221(1), pp. 144–184.

[3] Bob Coecke, Fabrizio Genovese, Martha Lewis, Dan Marsden & Alex Toumi (2018): *Generalized relations in linguistics & cognition*. Theoretical Computer Science 752, pp. 104–115, doi:https://doi.org/10.1016/j.tcs.2018.03.008. Available at `https://www.sciencedirect.com/science/article/pii/S0304397518301476`. Quantum structures in computer science: language, semantics, retrieval.

[4] Brendan Fong & David I Spivak (2019): *Hypergraph categories*. Journal of Pure and Applied Algebra 223(11), pp. 4746–4777.

[5] Karl Friston & Stefan Kiebel (2009): *Predictive coding under the free-energy principle*. Philosophical transactions of the Royal Society B: Biological sciences 364(1521), pp. 1211–1221.

[6] Tobias Fritz (2020): *A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics*. Advances in Mathematics 370, p. 107239.

[7] Soichiro Fujii (2019): *A categorical approach to L-convexity*. arXiv preprint arXiv:1904.08413.

[8] Grigori L Litvinov (2007): *Maslov dequantization, idempotent and tropical mathematics: A brief introduction*. Journal of Mathematical Sciences 140, pp. 426–444.

[9] Dan Marsden & Fabrizio Genovese (2017): *Custom hypergraph categories via generalized relations*. arXiv preprint arXiv:1703.01204.

[10] Peter McCullagh (2018): *Tensor methods in statistics*. Courier Dover Publications.

[11] Roger D Peng (2018): *Advanced statistical computing*. Work in progress, p. 121.

[12] R Tyrrell Rockafellar (1997): *Convex Analysis*. 11, Princeton University Press.

[13] Dario Stein (2021): *Structural foundations for probabilistic programming languages*. University of Oxford.

[14] Dario Stein & Richard Samuelson (2023): *A Category for unifying Gaussian Probability and Nondeterminism*. To Appear In LIPIcs Leibniz International Proceedings in Informatics.

[15] Simon Willerton (2015): *The Legendre-Fenchel transform from a category theoretic perspective*. arXiv preprint arXiv:1501.03791.

[16] Krzysztof Zajkowski (2017): *A variational formula on the Cramér function of series of independent random variables*. Positivity 21(1), pp. 273–282.