Pareto Frontier Learning with Expensive Correlated Objectives Supplementary Material

Amar Shah AS793@CAM.AC.UK

Zoubin Ghahramani ZOUBIN@ENG.CAM.AC.UK

Machine Learning Group, Department of Engineering, University of Cambridge

In this document we include details which were left out in the main text.

Expected improvement in Pareto Volume

In the main text, we defined the notion of *increase in Pareto volume* from a new observation and later partition the relevant region into grid cells to compute this value. Here we derive the equivalence in the way we define this quantity and the expression we end up using.

$$\begin{aligned}
\mathrm{Vol}_{\boldsymbol{v}_{\mathrm{ref}}}\Big(\mathcal{P}\big(\mathcal{Y}\cup\boldsymbol{y}\big)\Big) - \mathrm{Vol}_{\boldsymbol{v}_{\mathrm{ref}}}\Big(\mathcal{P}\big(\mathcal{Y}\big)\Big) &= \int\limits_{\mathbb{R}^L} \mathbb{I}\big[\boldsymbol{o}\succeq\boldsymbol{v}_{\mathrm{ref}}\big]\Bigg[\prod_{\boldsymbol{u}\in\mathcal{P}(\mathcal{Y})}\mathbb{I}\big[\boldsymbol{u}\not\succeq\boldsymbol{o}\big] - \prod_{\boldsymbol{u}\in\mathcal{P}(\mathcal{Y}\cup\boldsymbol{y})}\mathbb{I}\big[\boldsymbol{u}\not\succeq\boldsymbol{o}\big]\Bigg]\,d\boldsymbol{o} \\
&= \sum_{C\in\mathcal{C}}\ \int\limits_{\mathbb{R}^L\in C}\Bigg[\prod_{\boldsymbol{u}\in\mathcal{P}(\mathcal{Y})}\mathbb{I}\big[\boldsymbol{u}\not\succeq\boldsymbol{o}\big] - \prod_{\boldsymbol{u}\in\mathcal{P}(\mathcal{Y}\cup\boldsymbol{y})}\mathbb{I}\big[\boldsymbol{u}\not\succeq\boldsymbol{o}\big]\Bigg]\,d\boldsymbol{o} \\
&= \sum_{C\in\mathcal{C}_{\mathrm{nd}}}\ \int\limits_{\mathbb{R}^L\in C}\Bigg[\prod_{\boldsymbol{u}\in\mathcal{P}(\mathcal{Y})}\mathbb{I}\big[\boldsymbol{u}\not\succeq\boldsymbol{o}\big] - \prod_{\boldsymbol{u}\in\mathcal{P}(\mathcal{Y}\cup\boldsymbol{y})}\mathbb{I}\big[\boldsymbol{u}\not\succeq\boldsymbol{o}\big]\Bigg]\,d\boldsymbol{o} \\
&= \sum_{C\in\mathcal{C}_{\mathrm{nd}}}\ \int\limits_{\mathbb{R}^L\in C} 1-\mathbb{I}\big[\boldsymbol{y}\not\succeq\boldsymbol{o}\big]\ d\boldsymbol{o} \\
&= \sum_{C\in\mathcal{C}_{\mathrm{nd}}} \mathrm{Vol}_{\boldsymbol{v}_C}\big(\{\boldsymbol{y}\}\big). \qquad (1)
\end{aligned}$$

The first equality comes from the definition of the Pareto volume. The second equality partitions the domain of the integral. The third equality comes from the fact that the integrand is 0 for every $C\notin\mathcal{C}_{\mathrm{nd}}$. The fourth equality comes from the fact that for $\boldsymbol{o}\in C\in\mathcal{C}_{\mathrm{nd}}$, $\mathbb{I}\big[\boldsymbol{u}\not\succeq\boldsymbol{o}\big]=1$ for each $\boldsymbol{u}\in\mathcal{P}(\mathcal{Y})\cup\mathcal{P}(\mathcal{Y}\cup\boldsymbol{y})\backslash\{\boldsymbol{y}\}$, by definition of $\mathcal{C}_{\mathrm{nd}}$. The final equality comes from the definition of hypervolume in the main text. Consequently the expected increase in Pareto hypevolume from a new observation at x after observing data $\mathcal{D}$ is

$$\begin{aligned}
\mathrm{EIPV}(x|\mathcal{D}) &= \sum_{C\in\mathcal{C}_{\mathrm{nd}}}\int\limits_{C}\mathrm{Vol}_{\boldsymbol{v}_C}\big(\{\boldsymbol{y}\}\big)p(\boldsymbol{y}|\mathcal{D})\ d\boldsymbol{y} \\
&= \sum_{C\in\mathcal{C}_{\mathrm{nd}}}\int\limits_{\boldsymbol{v}_C}^{\boldsymbol{w}_C}\prod_{l=1}^{L}\big(y_l-v_{C,l}\big)p(\boldsymbol{y}|\mathcal{D})\ d\boldsymbol{y}. \qquad (2)
\end{aligned}$$

Integral of Product of Multivariate Gaussian Densities

We make use of a Gaussian approximation which leads to an analytically tractable quantity. Here we derive the integral with more detail. Let $\mathcal{N}(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Sigma})$ and $\mathcal{N}(\boldsymbol{y};\boldsymbol{\lambda},\boldsymbol{\Theta})$ be multivariate Gaussian density functions on $\mathbb{R}^L$. Then

$$
\begin{aligned}
&\int_{-\infty}^{\infty} \mathcal{N}(\boldsymbol{y};\boldsymbol{\lambda},\boldsymbol{\Theta})\mathcal{N}(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Sigma})\,d\boldsymbol{y} \\
&\qquad = (2\pi)^{-L}|\boldsymbol{\Theta}|^{-\frac{1}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\Big(\big(\boldsymbol{y}-\boldsymbol{\lambda}\big)^\top \boldsymbol{\Theta}^{-1}\big(\boldsymbol{y}-\boldsymbol{\lambda}\big) + \big(\boldsymbol{y}-\boldsymbol{\mu}\big)^\top \boldsymbol{\Sigma}^{-1}\big(\boldsymbol{y}-\boldsymbol{\mu}\big)\Big)\right) \\
&\qquad = (2\pi)^{-L}|\boldsymbol{\Theta}|^{-\frac{1}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\Big(\boldsymbol{\lambda}^\top\boldsymbol{\Theta}^{-1}\boldsymbol{\lambda} + \boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\Big)\right) \\
&\qquad\quad \times \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\Big(\boldsymbol{y}^\top(\boldsymbol{\Theta}^{-1}+\boldsymbol{\Sigma}^{-1})\boldsymbol{y} + 2\,\boldsymbol{y}^\top(\boldsymbol{\Theta}^{-1}\boldsymbol{\lambda}+\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\Big)\right) d\boldsymbol{y} \\
&\qquad = (2\pi)^{-L}|\boldsymbol{\Theta}|^{-\frac{1}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\Big(\boldsymbol{\lambda}^\top\boldsymbol{\Theta}^{-1}\boldsymbol{\lambda} + \boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\Big)\right) \\
&\qquad\quad \times \int_{-\infty}^{\infty} (2\pi)^{\frac{L}{2}}|\boldsymbol{\Omega}|^{\frac{1}{2}} \exp\left(\frac{1}{2}\boldsymbol{\nu}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\nu}\right)\mathcal{N}(\boldsymbol{y};\boldsymbol{\nu},\boldsymbol{\Omega})\,d\boldsymbol{y} \\
&\qquad = \exp\left(-\frac{1}{2}\Big(\boldsymbol{\lambda}^\top\boldsymbol{\Theta}^{-1}\boldsymbol{\lambda} + \log|\boldsymbol{\Theta}|\Big) - \frac{1}{2}\Big(\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \log|\boldsymbol{\Sigma}|\Big)\right. \\
&\qquad\quad \left. + \frac{1}{2}\Big(\boldsymbol{\nu}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\nu} + \log|\boldsymbol{\Omega}|\Big) - \frac{L}{2}\log(2\pi)\right), \qquad (3)
\end{aligned}
$$

where $\boldsymbol{\Omega}^{-1} = \boldsymbol{\Theta}^{-1} + \boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Omega}^{-1}\boldsymbol{\nu} = \boldsymbol{\Theta}^{-1}\boldsymbol{\lambda} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$.

Synthetic Function Definitions

Here we explicitly define the synthetic functions which we experiment on in the main text.

`oka2` is defined for $x_1 \in [-\pi, \pi]$, $x_2, x_3 \in [-5, 5]$ as

$$
\begin{aligned}
f_1(\boldsymbol{x}) &= -x_1, \\
f_2(\boldsymbol{x}) &= -1 + \frac{1}{4\pi^2}(x_1+\pi)^2 - |x_2 - 5\cos(x_1)|^{\frac{1}{3}} - |x_3 - 5\sin(x_1)|^{\frac{1}{3}}.
\end{aligned}
$$

`vlmop3` is defined for $\boldsymbol{x} \in [-3, 3]^2$ as

$$
\begin{aligned}
f_1(\boldsymbol{x}) &= -0.5(x_1+x_2)^2 - \sin(x^2+y^2), \\
f_2(\boldsymbol{x}) &= -\frac{(3x-2y+4)^2}{8} - \frac{(x-y+1)^2}{27} - 15, \\
f_3(\boldsymbol{x}) &= -\frac{1}{x^2+y^2+1} + 1.1\exp(-x^2-y^2).
\end{aligned}
$$

`dtlz1a` is defined for $\boldsymbol{x} \in [0, 1]^6$ as

$$
\begin{aligned}
f_1(\boldsymbol{x}) &= -0.5x_1(1+g(\boldsymbol{x})), \\
f_2(\boldsymbol{x}) &= -0.5(1-x_1)(1+g(\boldsymbol{x})), \\
g(\boldsymbol{x}) &= 100\left[5 + \sum_{i=2}^{6}(x_i-0.5)^2 - \cos\big(2\pi(x_i-0.5)\big)\right].
\end{aligned}
$$