Multi-Target Optimisation via Bayesian Optimisation and Linear Programming - Supplementary Material

Alistair Shilton **Santu Rana** **Sunil Gupta** **Svetha Venkatesh**
Deakin University, Geelong, Australia,
Centre for Pattern Recognition and Data Analytics
{alistair.shilton,santu.rana,sunil.gupta,svetha.venkatesh}@deakin.edu.au

The following proofs relate to theorems presented in section 5 of the paper.

Theorem 1 (Non-triviality) *Let $\boldsymbol{\alpha}^t$ be the solution to the score-function optimisation problem (9). Then $\boldsymbol{\alpha}^t \neq \mathbf{0}$.*

Proof: The constraints in the score-function optimisation problem ensure that $g_t(\mathbf{y}) < 0 \; \forall \mathbf{y} \in \mathcal{Y}_{t-1}$. Moreover as $g_t(\mathbf{y}) = 1$ if $\boldsymbol{\alpha}^t = \mathbf{0}$ it may be seen that to satisfy the constraints of the score-function optimisation problem it must be true that $\boldsymbol{\alpha}^t \neq \mathbf{0}$. □

Theorem 2 (Margin Minimisation) *Let $\boldsymbol{\alpha}^t$ be the solution to the score-function optimisation problem (9). Let $\mathbb{Y}^{g_t\star}$ be the estimated Pareto front defined by g_t. The minimum distance between $\mathcal{Y}_{t-1}$ and the estimated Pareto front $\mathbb{Y}^{g_t\star}$ is zero:*

$$\min_{\mathbf{y}\in\mathcal{Y}_{t-1}, \mathbf{y}'\in\mathbb{Y}^{g_t\star}} \|\mathbf{y} - \mathbf{y}'\| = 0$$

Proof: Suppose the theorem is false. Then by definition of the estimated Pareto front $\mathbb{Y}^{g_t\star}$ and observational consistency it must be true that $g_t(\mathbf{y}) < 0$ $\forall \mathbf{y} \in \mathcal{Y}_{t-1}$. Let $\delta = \min_{\mathbf{y}\in\mathcal{Y}_{t-1}} g_t(\mathbf{y}) < 0$. Defining $\bar{\boldsymbol{\alpha}}^t = (1+\delta)^{-1}\boldsymbol{\alpha}^t$ we see that $\bar{\boldsymbol{\alpha}}^t$ is a feasible solution to the score-function optimisation problem, and moreover $\|\bar{\boldsymbol{\alpha}}^t\|_1 < \|\boldsymbol{\alpha}^t\|_1$, so $\boldsymbol{\alpha}^t$ cannot be the optimal solution, which is a contradiction. It follows that the theorem must be true. □

Theorem 3 (Sparsity) *Let $\boldsymbol{\alpha}^t$ be the solution to the score-function optimisation problem (9). Then $\alpha_i^t = 0$ $\forall i : \mathbf{y}_i \notin \mathcal{Y}_{t-1}^\star$ (ie. points not in the estimated Pareto front cannot be support vectors).*

Proof: Let $\boldsymbol{\alpha}^t$ be the optimal solution of the score function optimisation problem with the additional constraint $\alpha_i^t = 0 \; \forall i : \mathbf{y}_i \notin \mathcal{Y}_{t-1}^\star$. It follows from Theorem 1 that $\sum_j \alpha_j^t L(\mathbf{y}_i, \mathbf{y}_j) > \frac{1}{2} \; \forall i : \mathbf{y}_i \notin \mathcal{Y}_{t-1}^\star$. Hence we see that $\boldsymbol{\alpha}^t$ is also the optimal solution to the score-function optimisation problem without the additional constraint, proving the theorem. □

Theorem 4 (Heaviside Limit) *Let $\kappa = \kappa_\perp$, where*

$$\kappa_\perp(y) = \lim_{\eta\to\infty} \frac{1}{1+\exp(-\eta y)} = \frac{1}{2}\left(1 + \mathrm{sgn}\left(y\right)\right),$$

and $\nexists i \neq j : \mathbf{y}_i = \mathbf{y}_j$. Then $\alpha_i^t = 1 \; \forall i : \mathbf{y}_i \in \mathcal{Y}_{t-1}^\star$, $\alpha_i^t = 0$ otherwise (ie. in the limiting case the support vectors are precisely the estimated Pareto set).

Proof: By theorem 3 it may be seen that $\alpha_i^t = 0 \; \forall i : \mathbf{y}_i \notin \mathcal{Y}_{t-1}^\star$. Moreover to satisfy the constraints of the score-function problem it must be true that $\forall i : \mathbf{y}_i \in \mathcal{Y}_{t-1}^\star$:

$$\begin{array}{l} \sum_j \alpha_j^t L\left(\mathbf{y}_i, \mathbf{y}_j\right) = \alpha_i^t \kappa_\perp\left(0\right) + \ldots \\ \ldots + \sum\limits_{j\neq i:\mathbf{y}_j\in\mathcal{Y}_{t-1}^\star} \alpha_j^t \kappa_\perp\left(\min_q\left(y_{iq} - y_{jq}\right)\right) = \frac{1}{2}\alpha_i^t \geq \frac{1}{2} \end{array}$$

and hence $\alpha_i^t = 1 \; \forall i : \mathbf{y}_i \in \mathcal{Y}_{t-1}^\star$. □