1 MIP formulations

Problem $(\mathcal{P})$ is formulated as a Mixed Integer Programming (MIP) to be handled by the commercial solvers CPLEX, GUROBI and MOSEK. For the experiment of Section 4.1 where f and h are given by (18a)-(18b), we reformulate the problem as

$$\left\{\begin{array}{rl} \min & \frac{1}{2}\lVert\mathbf{y}-\mathbf{A}\mathbf{x}\rVert_2^2+\lambda\mathbf{1}^{\mathrm{T}}\mathbf{z} \\ \text{s.t.} & -M\mathbf{z}\leq\mathbf{x}\leq M\mathbf{z} \\ & \mathbf{x}\in\mathbf{R}^n,\ \mathbf{z}\in\{0,1\}^n \end{array}\right. \tag{1}$$

that only involves linear and quadratic expressions. It can thus be handled by any of the three MIP solvers considered. Besides, the table below provides the MIP formulation of problem $(\mathcal{P})$ corresponding to the instances involved in Section 4.2. We remind that since CPLEX and GUROBI can handle second-order expressions and constraints, they cannot address instances with the loss (20). Such instances can be handled using MOSEK via conic expressions.

	Penalty (22a)	Penalty (22b)
Loss (19)	$\left\{\begin{array}{rl} \min & \frac{1}{2}\lVert\mathbf{y}-\mathbf{A}\mathbf{x}\rVert_2^2+\lambda\mathbf{1}^{\mathrm{T}}\mathbf{z}+\alpha\mathbf{1}^{\mathrm{T}}\mathbf{s} \\ \text{s.t.} & \mathbf{x}\geq-\mathbf{s} \\ & \mathbf{x}\leq\mathbf{s} \\ & \mathbf{x}\geq-M\mathbf{z} \\ & \mathbf{x}\leq M\mathbf{z} \\ & \mathbf{x}\in\mathbf{R}^n,\ \mathbf{z}\in\{0,1\}^n,\ \mathbf{s}\in\mathbf{R}^n \end{array}\right.$	$\left\{\begin{array}{rl} \min & \frac{1}{2}\lVert\mathbf{y}-\mathbf{A}\mathbf{x}\rVert_2^2+\lambda\mathbf{1}^{\mathrm{T}}\mathbf{z}+\alpha\mathbf{1}^{\mathrm{T}}\mathbf{s} \\ \text{s.t.} & \mathbf{x}\odot\mathbf{x}\leq\mathbf{s}\odot\mathbf{z} \\ & \mathbf{x}\geq-M\mathbf{z} \\ & \mathbf{x}\leq M\mathbf{z} \\ & \mathbf{x}\in\mathbf{R}^n,\ \mathbf{z}\in\{0,1\}^n,\ \mathbf{s}\in\mathbf{R}^n \end{array}\right.$
Loss (20)	$\left\{\begin{array}{rl} \min & \mathbf{1}^{\mathrm{T}}\mathbf{u}+\lambda\mathbf{1}^{\mathrm{T}}\mathbf{z}+\alpha\mathbf{1}^{\mathrm{T}}\mathbf{s} \\ \text{s.t.} & \mathbf{1}\geq\mathbf{v}+\mathbf{w} \\ & \mathbf{u}\geq-\log(\mathbf{v})+\mathbf{y}\odot\mathbf{A}\mathbf{x} \\ & \mathbf{u}\geq-\log(\mathbf{w}) \\ & \mathbf{x}\geq-\mathbf{s} \\ & \mathbf{x}\leq\mathbf{s} \\ & \mathbf{x}\geq-M\mathbf{z} \\ & \mathbf{x}\leq M\mathbf{z} \\ & \mathbf{x}\in\mathbf{R}^n,\ \mathbf{z}\in\{0,1\}^n,\ \mathbf{s}\in\mathbf{R}^n \\ & \mathbf{u}\in\mathbf{R}^m,\ \mathbf{v}\in\mathbf{R}^m,\ \mathbf{w}\in\mathbf{R}^m \end{array}\right.$	$\left\{\begin{array}{rl} \min & \mathbf{1}^{\mathrm{T}}\mathbf{u}+\lambda\mathbf{1}^{\mathrm{T}}\mathbf{z}+\alpha\mathbf{1}^{\mathrm{T}}\mathbf{s} \\ \text{s.t.} & \mathbf{v}+\mathbf{w}\leq\mathbf{1} \\ & \mathbf{u}\geq-\log(\mathbf{v})+\mathbf{y}\odot\mathbf{A}\mathbf{x} \\ & \mathbf{u}\geq-\log(\mathbf{w}) \\ & \mathbf{x}\odot\mathbf{x}\leq\mathbf{s}\odot\mathbf{z} \\ & -M\mathbf{z}\leq\mathbf{x}\leq M\mathbf{z} \\ & \mathbf{x}\in\mathbf{R}^n,\ \mathbf{z}\in\{0,1\}^n,\ \mathbf{s}\in\mathbf{R}^n \\ & \mathbf{u}\in\mathbf{R}^m,\ \mathbf{v}\in\mathbf{R}^m,\ \mathbf{w}\in\mathbf{R}^m \end{array}\right.$
Loss (21)	$\left\{\begin{array}{rl} \min & \lVert\mathbf{w}\rVert_2^2+\lambda\mathbf{1}^{\mathrm{T}}\mathbf{z}+\alpha\mathbf{1}^{\mathrm{T}}\mathbf{s} \\ \text{s.t.} & \mathbf{w}\geq\mathbf{1}-\mathbf{y}\odot\mathbf{A}\mathbf{x} \\ & \mathbf{w}\geq\mathbf{0} \\ & \mathbf{x}\geq-\mathbf{s} \\ & \mathbf{x}\leq\mathbf{s} \\ & \mathbf{x}\geq-M\mathbf{z} \\ & \mathbf{x}\leq M\mathbf{z} \\ & \mathbf{x}\in\mathbf{R}^n,\ \mathbf{z}\in\{0,1\}^n \\ & \mathbf{s}\in\mathbf{R}^n,\mathbf{w}\in\mathbf{R}^n \end{array}\right.$	$\left\{\begin{array}{rl} \min & \lVert\mathbf{w}\rVert_2^2+\lambda\mathbf{1}^{\mathrm{T}}\mathbf{z}+\alpha\mathbf{1}^{\mathrm{T}}\mathbf{s} \\ \text{s.t.} & \mathbf{w}\geq\mathbf{1}-\mathbf{y}\odot\mathbf{A}\mathbf{x} \\ & \mathbf{w}\geq\mathbf{0} \\ & \mathbf{x}\odot\mathbf{x}\leq\mathbf{s}\odot\mathbf{z} \\ & \mathbf{x}\geq-M\mathbf{z} \\ & \mathbf{x}\leq M\mathbf{z} \\ & \mathbf{x}\in\mathbf{R}^n,\ \mathbf{z}\in\{0,1\}^n \\ & \mathbf{s}\in\mathbf{R}^n,\mathbf{w}\in\mathbf{R}^n \end{array}\right.$

Table 1: MIP formulations used Section 4. The vectorial inequalities as well as the function "log" are taken component-wise and $\odot$ denotes the Hadarmad product.