Mathematical Optimization Model for Scrum-based Software Development

Gemini AI Assistant

September 4, 2025

Contents

1 Sets (Entities)

These sets are derived from the `Entities.csv` file and represent the foundational components of the domain model.

- **Projects (P)**: The set of all projects.
- **Teams (T)**: The set of all development teams.
- **Workers (W)**: The set of all individual workers.
- **Features (F)**: The set of all features.
- **Skills (S)**: The set of all professional skills.
- **Sprints (SP)**: The set of all sprints.
- **User Stories (US)**: The set of all user stories.
- **Tasks (TSK)**: The set of all tasks, the smallest units of work.
- **Blockers (BL)**: The set of all potential blockers.
- **Stakeholders (SH)**: The set of all stakeholders.
- **Release Plans (REP)**: The set of all release plans.

2 Indices

These indices are used throughout the mathematical model to refer to specific elements within the sets defined above.

- $p \in P$: Index for the set of Projects.
- $t \in T$: Index for the set of Teams.
- $w \in W$: Index for the set of Workers.
- $f \in F$: Index for the set of Features.
- $s \in S$: Index for the set of Skills.
- $sp \in SP$: Index for the set of Sprints.
- $us \in US$: Index for the set of User Stories.
- $tsk \in TSK$: Index for the set of Tasks.
- $bl \in BL$: Index for the set of Blockers.
- $sh \in SH$: Index for the set of Stakeholders.
- $rep \in REP$: Index for the set of Release Plans.

3 Decision Variables

These variables represent the choices that the optimization model can make. They are derived from `DecisionVariables.csv`.

DV0 **assign_worker_to_task** ($X_{w,tsk}$):

$$X_{w,tsk} \in \{0,1\}, \quad \forall w \in W, \forall tsk \in TSK$$

(1 if worker w is assigned to task tsk, 0 otherwise)

DV1 **assign_user_story_to_sprint** ($Y_{us,sp}$):

$$Y_{us,sp} \in \{0,1\}, \quad \forall us \in US, \forall sp \in SP$$

(1 if user story us is assigned to sprint sp, 0 otherwise)

DV2 **assign_team_to_project** ($A_{t,p}$):

$$A_{t,p} \in \{0,1\}, \quad \forall t \in T, \forall p \in P$$

(1 if team t is assigned to project p, 0 otherwise)

DV3 **assign_feature_to_release** ($B_{f,rep}$):

$$B_{f,rep} \in \{0,1\}, \quad \forall f \in F, \forall rep \in REP$$

(1 if feature f is assigned to release rep, 0 otherwise)

DV6 **set_team_size** ($Size_t$):

$$Size_t \in \mathbb{Z}^+, \quad \forall t \in T$$

(The number of workers assigned to team t)

DV8 **set_feature_priority** (P_f):

$$P_f \in \{1, ..., 10\}, \quad \forall f \in F$$

(The priority level assigned to feature f)

4 Goals

The goals represent the objective function of the optimization model. The overall aim is to find a solution that performs best against a weighted sum of these individual objectives, derived from `Goals.csv`. Let W_g be the weight for goal g. The multi-objective function is:

$$\text{Optimize} \quad Z = \sum_g W_g \cdot Z_g$$

Parameters for Goals

- $StoryPoints_{us}$: Story points for user story us.
- $Budget_p$: Allocated budget for project p.
- $Priority_f$: Priority score for feature f.
- $Influence_{sh}$: Influence level of stakeholder sh.
- $Effort_{tsk}$: Effort required for task tsk.

Goal Formulations

G0 **maximize_completed_story_points**:

$$\text{maximize} \sum_{us \in US} \sum_{sp \in SP} StoryPoints_{us} \cdot Y_{us,sp}$$

G1 **minimize_project_budget_overrun**: We model this by minimizing total cost, where cost is a function of assigned workers. Let $Cost_w$ be the cost of worker w.

$$\text{minimize} \sum_{w \in W} \sum_{tsk \in TSK} Cost_w \cdot X_{w,tsk}$$

G2 **maximize_feature_priority_delivery**:

$$\text{maximize} \sum_{f \in F} P_f$$

(Where P_f is the decision variable for setting priority)

G7 **minimize_total_task_effort**:

$$\text{minimize} \sum_{tsk \in TSK} Effort_{tsk}$$

(This is minimizing the effort of *selected* tasks. Assuming task selection is implicit via worker assignment)

$$\text{minimize} \sum_{w \in W} \sum_{tsk \in TSK} Effort_{tsk} \cdot X_{w,tsk}$$

G8 **maximize_stakeholder_value**: Assumes a link between features and stakeholders. Let $Rel_{f,sh}$ be a parameter (1 if feature f is relevant to stakeholder sh).

$$\text{maximize} \sum_{f \in F} \sum_{rep \in REP} \sum_{sh \in SH} (Influence_{sh} \cdot Rel_{f,sh}) \cdot B_{f,rep}$$

5 Conditions

The conditions represent the constraints that a valid solution must adhere to. They are derived from `Conditions.csv`.

Parameters for Conditions

- $Availability_w$: The availability (e.g., in hours) of worker w.
- $HasSkill_{w,s}$: Binary parameter, 1 if worker w has skill s.
- $NeedsSkill_{tsk,s}$: Binary parameter, 1 if task tsk requires skill s.
- $Status_w$: The current status of worker w (e.g., 1 for active, 0 for inactive).

Condition Formulations

C0 **maximize_skill_match_on_tasks** (Soft Constraint): This is a preference. It can be modeled as a goal to be maximized within the objective function.

$$\text{maximize} \sum_{w \in W} \sum_{tsk \in TSK} \sum_{s \in S} (HasSkill_{w,s} \cdot NeedsSkill_{tsk,s}) \cdot X_{w,tsk}$$

C1 **minimize_worker_overload** (Hard Constraint):

$$\sum_{tsk \in TSK} Effort_{tsk} \cdot X_{w,tsk} \leq Availability_w, \quad \forall w \in W$$

C2 **maximize_team_size_adherence** (Hard Constraint):

$$3 \leq Size_t \leq 9, \quad \forall t \in T$$

C5 **minimize_tasks_without_user_story** (Hard Constraint): Every assigned task must belong to an assigned user story. Let $BelongsTo_{tsk,us}$ be a binary parameter.

$$X_{w,tsk} \cdot BelongsTo_{tsk,us} \leq Y_{us,sp} \quad \forall w, tsk, us, sp$$

(A simpler formulation is to ensure each task is assigned to exactly one worker).

$$\sum_{w \in W} X_{w,tsk} \leq 1, \quad \forall tsk \in TSK$$

C6 **maximize_work_for_active_workers** (Hard Constraint):

$$\sum_{tsk \in TSK} X_{w,tsk} \leq M \cdot Status_w, \quad \forall w \in W$$

(Where M is a large number. If $Status_w$ is 0, the sum of assignments for worker w must be 0.)