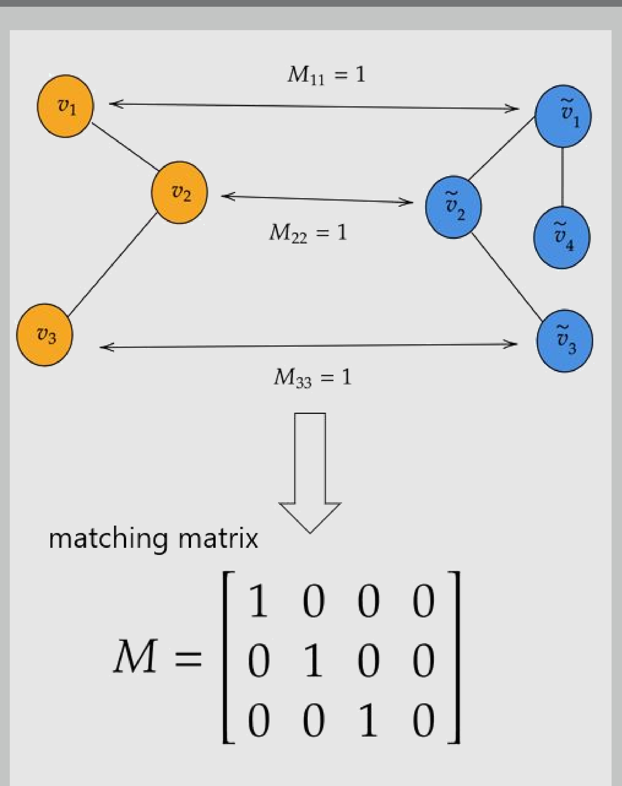
FRAM: Frobenius-Regularized Assignment Matching with Mixed-Precision Computing

Binrui Shen[1] Yuan Liang[1] Shengxin Zhu[1,2]

Beijing Normal University[1] Beijing Normal-Hong Kong Baptist University[2]

What is Graph Matching



Example

Objective Function and Method

The objective function of graph matching

$$\mathcal{Z}(M)=\frac{1}{2}\operatorname{tr}\left(M^T A M \tilde{A}\right)+\lambda \operatorname{tr}\left(M^T K\right).$$

The projected fixed-point method is

$$N^{(t)}=(1-\alpha)N^{(t-1)}+\alpha D^{(t)},$$
$$D^{(t)}=\mathcal{P}(\nabla \mathcal{Z}(N^{(t-1)})).$$

$\mathcal{P}(\cdot)$ projects the gradient onto a feasible set, such as a convex hull of the original domain, using the doubly stochastic projection (DSP):

$$\mathcal{P}_{\mathcal{D}}(X)=\arg\min_{D\in\mathcal{D}_{n\times n}}\|D-X\|_F.$$
$$\mathcal{D}_{n\times n}:=\{Y: Y\mathbf{1}=\mathbf{1}, Y^T\mathbf{1}=\mathbf{1}, Y\geq 0\}.$$

Challenges

Loss of scale invariance property

$$\arg\max \mathcal{Z}(M)=\arg\max w\mathcal{Z}(M),\quad w>0.$$
$$\mathcal{P}_{\mathcal{D}}(X)\neq\mathcal{P}_{\mathcal{D}}(wX),\quad X\in\mathbb{R}_+^{n\times n}.$$

Projection-relaxation may cause deviation from original domain.

Why DSP is Sensitive to Scaling Changes?

$$\mathcal{P}_{\mathcal{D}}(wX)=\arg\max_{D\in\mathcal{D}_{n\times n}}\langle D,X\rangle-\frac{1}{2w}\langle D,D\rangle.$$

Referring back to the updating formula, X corresponds to the gradient $\nabla\Phi(N^{(t)})$:

$$\mathcal{P}_{\mathcal{D}}(w\nabla\Phi(N^{(t)}))=\arg\max_{D\in\mathcal{D}_{n\times n}}-\frac{1}{2w}\langle D,D\rangle+\operatorname{tr}(D^T A N^{(t)}\tilde{A})+\lambda\operatorname{tr}(D^T K)$$

As w increases, the projection process emphasizes optimizing the objective assignment score.

Frobenius-Regularized Assignment (FRA)

Turn the uncontrollable problem's scale constant w into a controllable modeling parameter θ.

Theorem. DSP with a parameter θ

$$D_X^{\theta}=\arg\min_{D\in\mathcal{D}_{n\times n}}\|D-\frac{\theta}{2}X\|_F^2$$

can be rewritten as

$$D_X^{\theta}=\arg\max_{D\in\mathcal{D}_{n\times n}}\langle D,X\rangle-\frac{1}{\theta}\langle D,D\rangle.$$

The new formulation allows input normalization by $\frac{X}{\max(X)}$ to eliminate scaling variations (i.e., w).

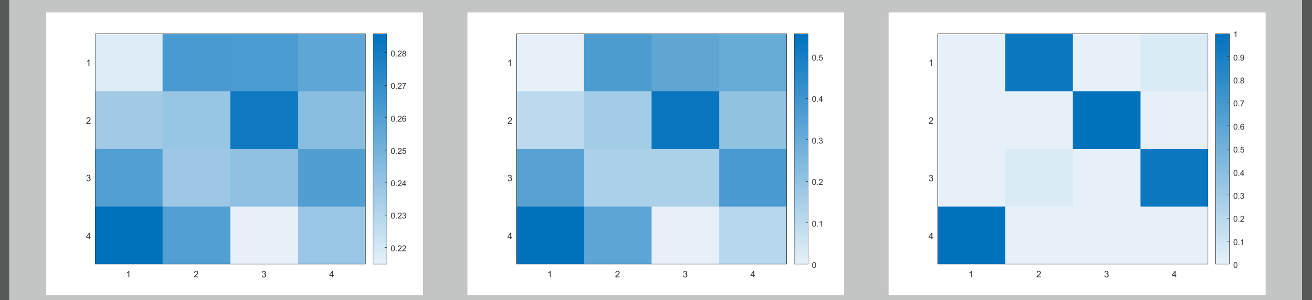
Effect of θ

What is the distance between D_X^{θ} and D_X^{∞}?

Proposition. For a matrix $X\in\mathbb{R}_+^{n\times n}$,

$$\frac{1}{n}\left(\langle D_X^{\infty},X\rangle-\langle D_X^{\theta},X\rangle\right)\leq\frac{1}{\theta}.$$

How θ suppresses the bias introduced by relaxation?



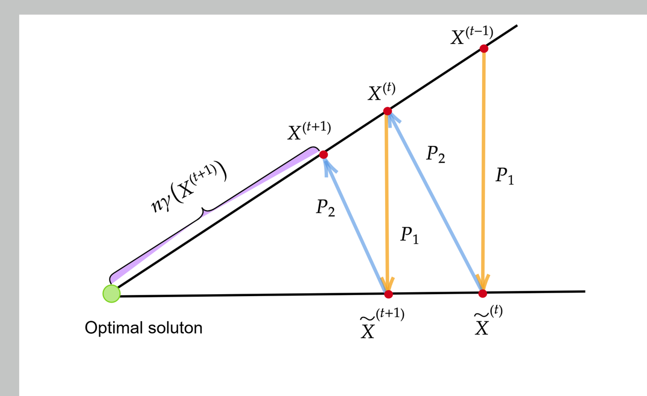
(a) $\theta=0.1$ (b) $\theta=1$ (c) $\theta=10$

Figure: Visualization of D_X^{θ} under different θ. The color of each cell represents the matrix entry, with darker shades indicating larger values.

Solve FRA

Scaling Doubly Stochastic Normalization (SDSN):

$$X^{(0)}=\frac{\theta X}{\max(X)},\quad \tilde{X}^{(k)}=\mathcal{P}_1\left(X^{(k-1)}\right),\quad X^{(k)}=\mathcal{P}_2\left(\tilde{X}^{(k)}\right),$$
$$\mathcal{P}_1(X)=\arg\min_{Y\mathbf{1}=Y^T\mathbf{1}=\mathbf{1}}\|X-Y\|_F,\ \mathcal{P}_2(X)=\arg\min_{Y\geq 0}\|X-Y\|_F^2.$$



Robustness to Rounding Error

Theorem Given that $X_k=\hat{X}_k+\Delta X_k$ is the k-th iterate of SDSN and ΔX_k is the rounding error, both X_k and $\hat{X}_k$ have the same limit under subsequent SDSN iterations.

This property prevents accumulation of rounding error arising from the gradient matrix computation and SDSN (corresponding to ΔX_0 and $\{\Delta X_k\}_{k>1}$).

Algorithm with Mixed-Precision

Algorithm FRAM

1: **Require** $A,\tilde{A},K,\lambda,\alpha,\theta,\delta_{th}$
2: Initial $X^{(0)}=\mathbf{0}_{n\times\tilde{n}}$
3: $c=\max(A,\tilde{A},K)$ ▷ FP64
4: $A=A/\sqrt{c},\ \tilde{A}=\tilde{A}/\sqrt{c},\ K=K/\sqrt{c}$ ▷ FP64
5: **While** $\delta^{(t)}>\delta_{th}$
6: $X^{(t)}=AN^{(t-1)}\tilde{A}+\lambda K$ ▷ TF32
7: $D^{(t)}=\mathrm{SDSN}(X^{(t)},\theta)$ ▷ FP32
8: $N^{(t)}=(1-\alpha)N^{(t-1)}+\alpha D^{(t)}$ ▷ FP64
9: $\delta^{(t)}=\|N^{(t)}-N^{(t-1)}\|_F/\|N^{(t)}\|_F$ ▷ FP64
10: *Discretize N to obtain M* ▷ FP64
11: **Return** Matching matrix M

Experiments on Facebook Networks

Social network	5% noise		15% noise		25% noise	
Methods	Acc	Time	Acc	Time	Acc	Time
S-GWL	26.4%	1204.1s	18.3%	1268.2s	17.9%	1295.8s
GWL	78.1%	3721.6s	68.4%	4271.3s	60.8%	4453.9s
DSPFP	79.7%	151.3s	68.3%	154.2s	62.2%	156.9s
GA	35.5%	793.2s	21.4%	761.7s	16.0%	832.6s
GRASP	37.9%	**63.6s**	20.3%	**67.4s**	15.7%	**71.3s**
ASM	91.1%	387.2s	88.4%	391.7s	85.7%	393.1s
AIPFP	68.6%	2705.5s	55.1%	2552.7s	47.8%	2513.8s
FRAM	**94.7%**	211.1s	**91.1%**	221.6s	**89.5%**	222.9s

Acceleration of Mixed-Precision Design

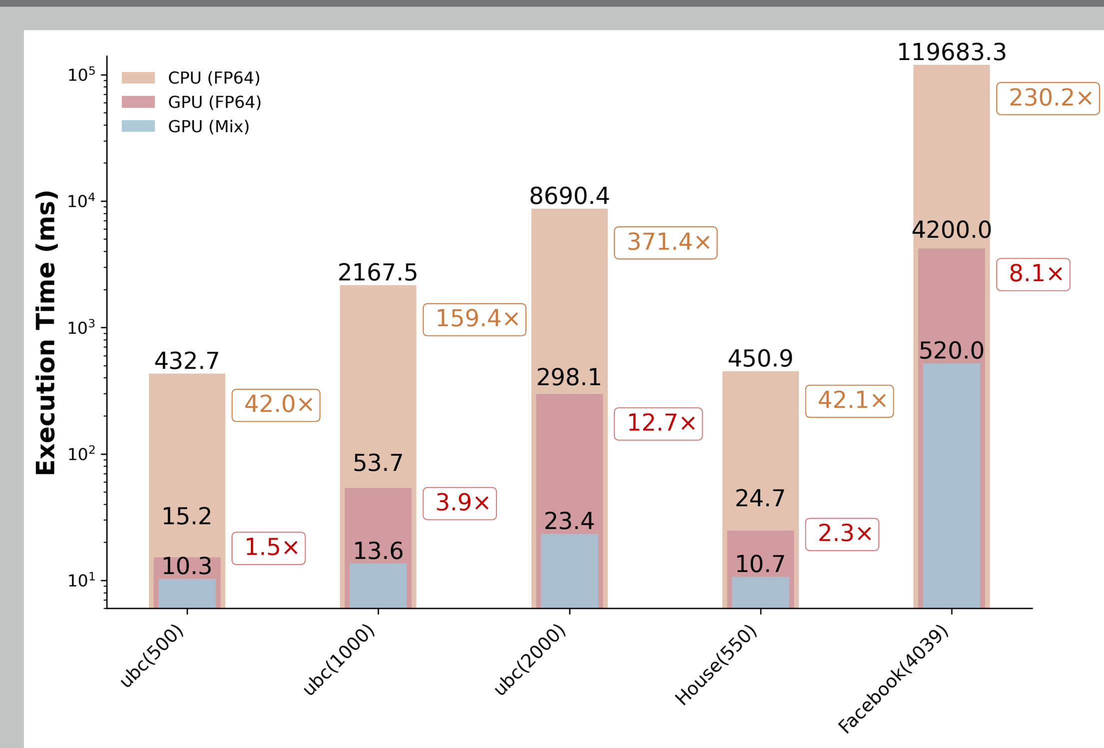


Figure: Red boxed numbers show the speedup of GPU mixed-precision over GPU double precision, and yellow boxed numbers show the speedup over CPU double precision.