Analysis and Generalization of the HiPPI Algorithm for Multi-Graph Matching

M. Kahl and B. Savchynskyy

1. Introduction

Optimally assigning elements of finite sets to each other is a fundamental combinatorial problem occurring in a plethora of applications ranging from economics to computer vision. The notion of optimality is defined with respect to a cost tensor, whose dimensionality predominantly determines the problem's computational complexity. Relating one element per set results in the polynomially solvable linear assignment problem (LAP), couples of elements, conversely, in the NP-hard quadratic assignment problem (QAP). In order to match multiple sets, several assignment problems are coupled via cycle-consistency constraints - giving rise to Multi-LAPs or -QAPs. In computer vision, applications usually demand distance preserving matchings of various visual objects, which on the contrary necessitate solving Multi-QAPs. Video-based tracking, shape modelling or multi-view reconstruction are just some of a wide range of examples in this context. In [1], Bernard et al. set out to tackle problems of this nature by jointly addressing both - cost construction and solution of the related assignment problem. Dissecting their approach and its culmination, higher-order projected power iterations (HiPPI), is the subject of this report.

Although this report mainly acts as a review of Bernard et al.'s work, it contributes in a threefold way: i) Bernard et al.'s problem formulation is placed in a broader context - highlighting its shortcomings, ii) a new convergence proof is introduced - revealing a novel perspective on HiPPI's operating principle, iii) HiPPI is generalized to arbitrary Multi-QAP cost structures.

2. Background

Pairwise Matching concerns itself with the matching of two sets $\mathcal{V}^{[1]}, \mathcal{V}^{[2]}$, such that $n_p := |\mathcal{V}^{[p]}| < \infty$, $p = 1, 2$. Allowing for unassigned elements, feasible matchings are encoded by partial permutation matrices $\mathbb{P}_{m,n} := \left\{X \in \{0,1\}^{m\times n} \mid X\vec{1}_n \le \vec{1}_m, X^T\vec{1}_m \le \vec{1}_n, \right\}$, where we introduce the notation $\mathbb{P}^{[p,q]} := \mathbb{P}_{n_p,n_q}$. For $X \in \mathbb{P}^{[1,2]}$, $X_{i_1 i_2} = 1$ is consequently interpreted as assigning $i_1 \in \mathcal{V}^{[1]}$ to $i_2 \in \mathcal{V}^{[2]}$. Minimizing a cost function $C : \mathcal{V}^{[1]} \times \mathcal{V}^{[2]} \to \bar{\mathbb{R}}$ over the set of feasible matchings gives rise to the linear assignment problem [2, Sec. 1.2]

$$\min_{X\in\mathbb{P}^{[1,2]}} \left\{ \langle \vec{C}, \vec{X} \rangle = \langle C, X \rangle_F = \sum_{(i_1,i_2)\in\mathcal{V}^{[1]}\times\mathcal{V}^{[2]}} C_{i_1 i_2} X_{i_1 i_2} \right\}. \tag{1}$$

Here, $\langle\cdot,\cdot\rangle$ denotes the inner product of the respective Euclidean space, $\langle\cdot,\cdot\rangle_F$ the Frobenius inner product with the canonical identification $C \cong [C(i_1,i_2)]_{i_1,i_2} \in \overline{\mathbb{R}}^{n_1\times n_2}$, and $\vec{X} := \text{vec}(X)$ the column-wise vectorization of a matrix. LAPs are solvable in polynomial time e. g. via the Hungarian Algorithm [3] with a complexity of $O(\overline{n}^3)$ [4] or the Auction Algorithm [5] with an (empirical) average time complexity of $O(\overline{n}^2\log(\overline{n}))$ [6], where $\overline{n} := \max\{n_1, n_2\}$. In [1], Bernard et al. rely on an implementation of the Auction Algorithm specified in [7].

Including pairwise costs via a higher-order cost function $C: \left(\mathcal{V}^{[1]}\times\mathcal{V}^{[2]}\right)^2 \to \overline{\mathbb{R}}$ results, using the identification $C \cong \left[C(i_1,i_2,j_1,j_2)\right]_{i_1 i_2, j_1 j_2} \in \overline{\mathbb{R}}^{n_1 n_2\times n_1 n_2}$, in the quadratic assignment problem [8]

$$\min_{X\in\mathbb{P}^{[1,2]}} \left\{ \langle \vec{X}, C\vec{X} \rangle = \sum_{\substack{(i_1,i_2)\in\mathcal{V}^{[1]}\times\mathcal{V}^{[2]} \\ (j_1,j_2)\in\mathcal{V}^{[1]}\times\mathcal{V}^{[2]}}} C_{i_1 i_2, j_1 j_2} X_{i_1 i_2} X_{j_1 j_2} \right\}. \tag{2}$$

Higher-order generalizations for cost functions $C: \left(\mathcal{V}^{[1]}\times\mathcal{V}^{[2]}\right)^N \to \overline{\mathbb{R}}$ are treated in [9], [10], but are uninteresting for this report. A LAP with costs C_{LAP} can be formulated as an equivalent QAP using costs $C_{\text{QAP}} := \text{diag}\left(\vec{C}_{\text{LAP}}\right)$, showing that the latter is a strict generalization of the former. The QAP in Eq. (2) is given in Lawler's form [8], which will hereafter just be referred to as QAP. Parametrizing costs as $C = D\otimes F + \text{diag}\left(\vec{B}\right)$ via $F\in\overline{\mathbb{R}}^{n_1\times n_1}$, $D\in\overline{\mathbb{R}}^{n_2\times n_2}$, and $B\in\overline{\mathbb{R}}^{n_1\times n_2}$ yields the QAP in Koopmanns-Beckman's form [11, Sec. 6]

$$\min_{X\in\mathbb{P}^{[1,2]}} \left\{ \langle X, FXD^T + B\rangle_F = \sum_{\substack{(i_1,i_2)\in\mathcal{V}^{[1]}\times\mathcal{V}^{[2]} \\ (j_1,j_2)\in\mathcal{V}^{[1]}\times\mathcal{V}^{[2]}}} F_{i_1 j_1} D_{i_2 j_2} X_{i_1 i_2} X_{j_1 j_2} + \sum_{(i_1,i_2)\in\mathcal{V}^{[1]}\times\mathcal{V}^{[2]}} B_{i_1 i_2} X_{i_1 i_2} \right\}. \tag{3}$$

QAPs are in general strongly NP-hard [12, Vol. 5, Sec. 3.4, Thm. 2]. If each element is assigned exactly once, one speaks of a complete QAP, in contrast to the up to here considered incomplete QAPs. In this case one optimizes over the set of complete permutation matrices. In [13], Haller et al. polynomially reduce the incomplete to the complete QAP.

Multi-Matching treats the matching of $k\in\mathbb{N}$ sets $\mathcal{V}^{[p]}$ with $n_p := |\mathcal{V}^{[p]}| < \infty$, where $p\in[\![k]\!] := [0,k]\cap\mathbb{N}$. Multiple matchings $X^{[p,q]}\in\mathbb{P}^{[p,q]}$ between pairs of sets $\mathcal{V}^{[p]}, \mathcal{V}^{[q]}$, $p,q\in[\![k]\!]$ allow for inconsistencies - a disagreement between the first and last element of a chain obtained by traversing matchings along a cycle of sets. To avoid this scenario one enforces cycle-consistency for all 3-cycles, which is sufficient for global consistency [14, Prop. 1], see Def. 1.

Definition 1. *(Cycle-Consistency, [1, Def. 1])*
A multi-matching $X = \left[X^{[p,q]}\right]_{(p,q)\in[\![k]\!]^2} \in \left[\mathbb{P}^{[p,q]}\right]_{(p,q)\in[\![k]\!]^2}$ is said to be cycle-consistent if

$$\forall p,q,r\in[\![k]\!]: \begin{cases} \textit{i)} & X^{[p,p]} = I_{n_p} & \textit{(identity)} \\ \textit{ii)} & \left(X^{[p,q]}\right)^T = X^{[q,p]} & \textit{(symmetry)} \\ \textit{iii)} & X^{[p,r]}X^{[r,q]} \leq X^{[p,q]} & \textit{(transitivity)}. \end{cases} \tag{4}$$

The set of cycle-consistent multi-matchings is denoted with $\mathbb{X}$.

Cycle-consistency can alternatively be characterized by means of the universe concept. Matching each set $\mathcal{V}^{[p]}$ solely to a set of $d \in \mathbb{N}$ universe points $[\![d]\!]$, while considering elements matched to the same universe point as matched, results in a cycle-consistent matching. Lemma 1 concretizes this idea.

Lemma 1. *(Universe Characterization of Cycle-Consistency)*
Assume $X \in \left[\mathbb{P}^{[p,q]}\right]_{(p,q)\in[\![k]\!]^2}$. Then $(X \in \mathbb{X}) \quad \Leftrightarrow \quad \left(\exists d \in \mathbb{N} : \exists U \in \mathbb{U}_d : \forall p,q \in [\![k]\!] : X^{[p,q]} = U^{[p]}\left(U^{[q]}\right)^T\right)$, where $\mathbb{U}_d := \left\{\left[U^{[p]}\right]_{p\in[\![k]\!]} \in \left[\mathbb{P}_{n_p,d}\right]_{p\in[\![k]\!]} \middle| \forall p \in [\![k]\!] : U^{[p]}\vec{1}_d = \vec{1}_{n_p}\right\}$ denotes the set of universe matchings for a universe of size d. Rephrased, it holds that $X = UU^T$.

Proof. For "$\Leftarrow$", see [1, Lemma 2] and for "$\Rightarrow$", see A.1. □

Optimizing over $\mathbb{U}_d$ instead of $\mathbb{X}$ by means of the substitution from Lemma 1 comes, on one hand, at the expense of more variables and an additional parameter in form of the universe size d, but has, on the other hand, the benefit of much simpler constraints.
Now, Multi-Assignment-Problems (MAP) of order N relate N cycle-consistent matchings between up to $2N$ sets $\mathcal{V}^{[p_i]}$, $i \in [\![2N]\!]$. More specifically, a MAP of order 2, or Multi-QAP, tries to optimize cost-functions $C^{[p,q,r,s]} : \mathcal{V}^{[p]} \times \mathcal{V}^{[q]} \times \mathcal{V}^{[r]} \times \mathcal{V}^{[s]} \mapsto \bar{\mathbb{R}}$ over the set of cycle-consistent multi-matchings,

$$\min_{X\in\mathbb{X}} \left\{ \sum_{(p,q,r,s)\in[\![k]\!]^4} \langle \vec{X}^{[p,q]}, C^{[p,q,r,s]} \vec{X}^{[r,s]} \rangle \right\}, \tag{5}$$

where again the identification $C^{[p,q,r,s]} \cong \left[C^{[p,q,r,s]}(i_p, i_q, i_r, i_s)\right]_{i_p i_q, i_r i_s} \in \bar{\mathbb{R}}^{n_p n_q \times n_r n_s}$ is made.
A frequently encountered special case is the Multi-QAP with pairwise decomposable costs. In line with [15, Sec. III, A.], only costs $C^{[p,q]} : \left(\mathcal{V}^{[p]} \times \mathcal{V}^{[q]}\right)^2 \to \bar{\mathbb{R}}$ between pairs of sets are established. Using $C^{[p,q,r,s]} := \left\{ \begin{array}{ll} C^{[p,q]}; & r = p \wedge q = s \\ 0; & else \end{array} \right\}$ one obtains the optimization problem,

$$\min_{X\in\mathbb{X}} \left\{ \sum_{(p,q)\in[\![k]\!]^2} \langle \vec{X}^{[p,q]}, C^{[p,q]} \vec{X}^{[p,q]} \rangle \right\}. \tag{6}$$

MAPs of different orders, such as the Multi-LAP, are defined in analogy to Eq. (5). Note, however, as shown in [16, Thm. 1], that, in contrast to the LAP, the Multi-LAP is NP-hard.

3. Related Work & Contribution

Two major categories exist to classify methods solving MAPs: i) permutation synchronization e. g. [17, 18, 19] and ii) joint optimization e. g. [10, 20, 21]. Permutation synchronization focuses, as the name might indicate, on retrieving a cycle-consistent from a non-cycle-consistent multi-matching. Usually only pairs of objects are considered to obtain the non-cycle-consistent multi-matching, making the approach fairly scalable. The most prominent drawback of the synchronization step, often just designed as a projection onto the set of feasible multi-matchings, is its ignorance w.r.t. costs, e. g. the ones used to obtain pairwise matchings. Joint optimization methods, on the contrary, try to optimize over the set of feasible multi-matchings, likely generating higher quality solutions, but, at the same time, being far more expensive than synchronization approaches. Both categories, essentially relying on permutation matrices, allow for various relaxations, e. g. semidefinite programming relaxations [22, 23], spectral relaxations [17, 19], or low-rank relaxations [24] - with some of these, cycle-consistency is not always guaranteed.

Almost as diverse as the approaches to optimize MAPs are the approaches to model them. The primary question is still how to obtain costs and if these should be merely linear or quadratic, but with computer vision applications emphasizing geometric consistency, more creative formulations have emerged, [25], for example, enforce the low rank of 2D coordinate matrices of orthographic projections of 3D scenes as an additional constraint.

In HiPPI Bernard et al. place a novel method into this landscape, that jointly optimizes for multi-matchings, incorporates geometric relations, guarantees cycle-consistency, and scales well (in light of their experiments, see Sec. 5).

4. Higher-Order Projected Power Iterations

4.1. Problem Formulation

The Multi-QAP formulation used in [1] by Bernard et al. is based on universe matchings to a universe of priorly fixed size d and weights $W = \left[W^{[p,q]}\right]_{(p,q)\in[\![k]\!]^2} \in \mathbb{S}_n^+$, where $\mathbb{S}_n^+ := \left\{M \in \mathbb{R}^{n\times n} \,\middle|\, M \text{ symmetric, positive semidefinite}\right\}$, $n := \sum_{p\in[\![k]\!]} n_p$, and reads

$$\max_{U\in\mathbb{U}_d} \left\{\langle U^T W U, U^T W U\rangle_F =: f(U)\right\}. \tag{7}$$

4.1.1. Motivation

Motivated by computer vision applications Bernard et al. firstly assume their elements $i_p \in \mathcal{V}^{[p]}$ to be embedded in a metric space $(\mathcal{X}, d)$, e. g. 2D image coordinates, and secondly to be equipped with Euclidean, f-dimensional feature vectors $F_{i_p} \in \mathbb{R}^f$. With these assumptions adjacency matrices for all sets $A^{[p]} := \left[\exp\left(-d(i_p, j_p)^2\right)\right]_{(i_p,j_p)\in(\mathcal{V}^{[p]})^2} \in \mathbb{S}_{n_p}^+$, summarized in $A := \operatorname{diag}\left(A^{[p]}, p \in [\![k]\!]\right) \in \mathbb{S}_n^+$, and similarity matrices for pairs of sets $S^{[p,q]} := \left[\exp\left(-\|F_{i_p} - F_{i_q}\|^2\right)\right]_{(i_p,i_q)\in\mathcal{V}^{[p]}\times\mathcal{V}^{[q]}} \in \mathbb{R}_+^{n_p\times n_q}$, summarized in $S := \left[S^{[p,q]}\right]_{(p,q)\in[\![k]\!]^2}$, are used to define weighted adjacency matrices $W := S^T A S \in \mathbb{S}_n^+$. Scale and bandwidth of the Gaussian kernels were subject to application dependent adaption. Interpreting terms of the form $\langle (U^{[p]})^T A^{[p]} U^{[p]}, (U^{[q]})^T A^{[q]} U^{[q]} \rangle$ as agreement of the matrices $A^{[p]}$ and $A^{[q]}$ after reordering both according to the universe and SU as reweighted universe matchings, that emphasize matchings with high similarity, mediates at least some intuition regarding the problem formulation. At this point it should be pointed out, that, although all in [1] presented applications made use of such a structure, only the requirement $W \in \mathbb{S}_n^+$ is necessary for the main results of the paper. Particularly, any kernel could be used to construct A - as long as $A \in \mathbb{S}_n^+$, $W \in \mathbb{S}_n^+$ holds.

4.1.2. Substitution of Universe Matchings and Reduction to Pairwise Matching

Details are for brevity's sake omitted and can be found in Sec. A.2.
Substituting $X = UU^T$ in light of Lemma 1 shows that the formulation in Eq. (7) is not as general as the Multi-QAP (5), but more general than the Multi-QAP with pairwise decomposable costs (6),

$$f(U) = \langle WXW, X \rangle_F = \sum_{(p,q,r,s)\in [\![k]\!]^4} \langle \vec{X}^{[r,s]}, \left(W^{[q,s]}\right)^T \otimes W^{[r,p]} \vec{X}^{[p,q]} \rangle = f(X)\,. \tag{8}$$

The previous equation engages in the hereafter often committed abuse of notation of reusing objects defined w.r.t. either universe or multi-matchings in terms of the other, e. g. $f(U) = f(X)$.

Reducing Eq. (7) to $k = 2$ reveals a problem, that again floats midst two formulations, now Lawler's form (2) and Koopmanns-Beckman's form (3),

$$\begin{aligned} f(X) = \quad & \langle W^{[1,2]}, W^{[1,2]} \rangle_F \\ + \quad & 4 \left\{ \textstyle\sum_{(i_1 i_2) \in \mathcal{V}^{[1]} \times \mathcal{V}^{[2]}} \left(\sum_{p \in [\![2]\!]} \sum_{i_p \in \mathcal{V}^{[p]}} W^{[1,p]}_{i_1 i_p} W^{[p,2]}_{i_p i_2} \right) \quad X^{[1,2]}_{i_1 i_2} \right\} \\ + \quad & 4 \left\{ \textstyle\sum_{(i_1, i_2, j_1, j_2) \in (\mathcal{V}^{[1]} \times \mathcal{V}^{[2]})^2} \left(W^{[1,1]}_{i_1 j_1} W^{[2,2]}_{i_2 j_2} + W^{[1,2]}_{j_1 i_2} W^{[1,2]}_{i_1 j_2} \right) \quad X^{[1,2]}_{i_1 i_2} X^{[1,2]}_{j_1 j_2} \right\}. \end{aligned} \tag{9}$$

Considering the motivation, it is the reweighing of universe matchings, SU, that differentiates Eq. (7) from the simpler decomposable costs or Koopmanns-Beckman's forms. Bypassing this by setting $S = I_n$, and taking $W = S^T AS$ into account one obtains a Multi-QAP with decomposable costs, which are in turn all in Koopmanns-Beckman's form,

$$f(X) = \sum_{(p,q)\in [\![k]\!]^2} \langle \vec{X}^{[p,q]}, A^{[q]} \otimes A^{[p]} \vec{X}^{[p,q]} \rangle\,. \tag{10}$$

4.2. Iteration

The Algorithm deployed in [1] to tackle problem (7) requires an initial universe matching $U_{\{t=0\}} \in \mathbb{U}_d$ to a universe of priorly fixed size d and proceeds to perform the update

$$\begin{array}{lll} U_{\{t+1\}} & \leftarrow & \mathrm{proj}_{\mathbb{U}_d}\left(W U_{\{t\}} U_{\{t\}}^T W U_{\{t\}} \right) \qquad \text{(HiPPI)} \\ t & \leftarrow & t+1 \end{array}\,, \tag{11}$$

until $|f(U_{\{t\}}) - f(U_{\{t-1\}})| = 0$. Projections are, if not otherwise stated, meant as a mapping to the minimizer of an appropriate norm to the set in consideration.
Convergence is shown in [1, Corollary 5] or via an arguably simpler proof in Sec. 4.3. The initial universe matching $U_{\{0\}}$ has to be obtained via a different solver. Bernard et al. use Quickmatch [20], a clustering-based algorithm with no demand for any initialisation. Note, that this technically renders HiPPI a local search method. Almost as its name, HiPPI, suggests, the update can be dissected into four parts - linearization, relaxation, power iteration, and projection, all of which will be elaborated upon in the following sections. Beware that these sections possibly reuse some already defined symbols in a different context. This is done to preserve the common thread relating all these methods throughout the report.

4.2.1. Power Iteration

HiPPI in its most stripped down form is, as the acronym indicates, a power, or strictly speaking, an orthogonal iteration. Given a matrix $W \in \mathbb{R}^{n\times n}$ and $d \in [\![n]\!]$ orthonormal columns arranged in a matrix $U_{\{0\}} \in \mathbb{R}^{n\times d}$, the orthogonal iteration performs the update

$$U_{\{t+1\}}R \overset{\text{QR-decomposition}}{\hookleftarrow} WU_{\{t\}}\,. \tag{12}$$

The columns of $U_{\{t\}}$ converge to a basis of the eigenspace spanned by the eigenvectors of W with the largest d eigenvalues λ_i under the sensible assumption of its separation, $|\lambda_d| > |\lambda_{d+1}|$ (and additional, but mild conditions on $U_{\{0\}}$), as shown in [26, Thm. 8.2.2]. [27, Lemma 2] shows that such a basis is a solution to the generalized Rayleigh problem,

$$\max_{U\in\mathbb{ST}_{nd}} \{\langle U, WU\rangle_F\}\,, \tag{13}$$

whose feasible set constitutes the Stiefel manifold $\mathbb{ST}_{nd} := \left\{U \in \mathbb{R}^{n\times d} \,\middle|\, U^T U = I_d\right\}$. The connection to optimization on manifolds reveals a more generalizable perspective on the update of the orthogonal iteration than Eq. (12). Although the projection onto the Stiefel manifold is given by the polar decomposition [28, Thm. 8.4], which is well-defined, if $\operatorname{rank}(WU_{\{t\}}) = d \left(\Leftarrow W \in \mathbb{S}_n^+ \wedge \lambda_d > \lambda_{d+1}\right)$, it is linked to the QR-decomposition by the fact that both are retractions onto the Stiefel manifold in the sense of [29, Def. 4.1.1]. Conceptually speaking, retractions are gradient preserving maps from the tangent bundle to the manifold, figuratively (and loosely) speaking, they map iterates of a line search method back to the manifold once deviated. Briefly ignoring that update (12) does not necessarily perform a line-search on the Stiefel manifold, therefore retractions are possibly undefined, one could think of Eq. (12), in a very broad sense, as,

$$U_{\{t+1\}} \hookleftarrow \text{"proj"}_{\mathbb{ST}_{nd}}\left(WU_{\{t\}}\right), \tag{14}$$

which is already more reminiscent of Eq. (11).

4.2.2. Relaxation & Projection

Following the line of thought suggested by Eq. (14), one should replace the Stiefel manifold in problem (13) with the universe matchings $\mathbb{U}_d$ as the feasible set,

$$\max_{U \in \mathbb{U}_d} \{\langle U, WU \rangle_F\} . \tag{15}$$

The orthogonal iteration (12) addresses a relaxation of this problem. This is the conclusion of Lemmas 2 - 4, which require the 'generalized' Stiefel manifold (a protologism unused in this sense beyond this report) $\mathbb{GST}_{nd} := \{U \in \mathbb{R}^{n \times d} \,|\, U^T U \le I_d\}$. Proofs can be found in Sec. A.3.

Lemma 2. $\max_{U \in \sqrt{k}\mathbb{GST}_{nd}} \{\langle U, WU \rangle_F\}$ *is a relaxation of* $\max_{U \in \mathbb{U}_d} \{\langle U, WU \rangle_F\}$.

Lemma 3. $\operatorname*{arg\,max}_{U \in \sqrt{k}\mathbb{GST}_{nd}} \{\langle U, WU \rangle_F\} = \sqrt{k} \operatorname*{arg\,max}_{U \in \mathbb{GST}_{nd}} \{\langle U, WU \rangle_F\} = \sqrt{k} \operatorname*{arg\,max}_{U \in \mathbb{ST}_{nd}} \{\langle U, WU \rangle_F\}$, *where the second equality requires* $W \in \mathbb{S}_n^+$ *with eigenvalues* $\lambda_1, \ldots, \lambda_d > \lambda_{d+1}, \ldots, \lambda_n$.

Lemma 4. $\operatorname{proj}_{\mathbb{U}_d}\left(\sqrt{k} V\right) = \operatorname{proj}_{\mathbb{U}_d}(V)$

The relaxation of problem (15) from Lemma 2 is by means of Lemma 3 related to the generalized Rayleigh problem (13). This means, scaling a solution obtained via the orthogonal iteration (12) by $\sqrt{k}$ would address the relaxed problem. With projections back to the original feasible set $\mathbb{U}_d$ being common practice in similar scenarios, think of projected gradient descent or [30], this scaling can be skipped due to the scale invariance of the projection, seen in Lemma 4. Therefore, the relaxed problem is actually addressed by an orthogonal iteration, while a successive projection onto $\mathbb{U}_d$ addresses problem (15) - or in other words, problem (15) is addressed by the update,

$$U_{\{t+1\}} \leftarrow \operatorname{proj}_{\mathbb{U}_d}\left(WU_{\{t\}}\right), \tag{16}$$

which is not too dissimilar from the actual HiPPI update (11), just as problem (15) is not too dissimilar from problem (7) addressed by HiPPI. Before this relationship can be explored further, one should elucidate the projection onto the universe matchings $\mathbb{U}_d$. As $\mathbb{U}_d$ is a discrete non-singleton, therefore not convex, the projection is not necessarily well-defined. Nonetheless, the accompanying optimization problem turns out to be efficiently solvable as it is equivalent to k independent LAPs [27, Sec. 4.2.], which can be seen if one realizes $\mathbb{U}_d = \left[\bar{\mathbb{P}}_{n_p d}\right]_{p \in [\![k]\!]}$, $\bar{\mathbb{P}}_{kl} := \{P \in \mathbb{P}_{kl} \,|\, P\vec{1}_l = \vec{1}_k\}$ and observes,

$$\operatorname{proj}_{\mathbb{U}_d}(V) = \operatorname*{arg\,min}_{U \in \mathbb{U}_d} \left\{\|U - V\|_F^2\right\} = \operatorname*{arg\,min}_{U \in \mathbb{U}_d} \left\{\|V\|_F^2 + \|U\|_F^2 - 2\langle U, V \rangle_F\right\} \overset{\|U\|_F^2 = n}{=} \operatorname*{arg\,max}_{U \in \mathbb{U}_d} \{\langle U, V \rangle_F\} . \tag{17}$$

4.2.3. Linearization

For subsequent discussion it is useful to first fix further notation. The linearization of a differentiable, scalar function $g : C \to \mathbb{R}$ at $x_0 \in C$ is denoted by $g_{x_0}(x) := g(x_0) + \partial g(x_0)(x - x_0) \sim \partial g(x_0)(x)$, where $\sim$ symbolizes the dropping of, for optimization purposes usually uninteresting, constant or positive scaling factors. Detailed calculations can be found in Sec. A.4. In line with the abuse of notation from Eq. (8), we use $X_{\{t\}} = U_{\{t\}} U_{\{t\}}^T$. Taking a look at the reformulated objective $f(X)$ and its linearization we find,

$$f_{X_{\{t\}}}(X) \sim \langle \underbrace{W X_{\{t\}} W}_{:= W_{\{t\}}}, X \rangle_F = \langle U, W_{\{t\}} U \rangle_F \sim f_{X_{\{t\}}}(U), \tag{18}$$

which brings us finally back to the HiPPI update (11), now using $W_{\{t\}}$,

$$U_{\{t+1\}} \leftarrow \text{proj}_{\mathbb{U}_d} \left(W U_{\{t\}} U_{\{t\}}^T W U_{\{t\}} \right) = \text{proj}_{\mathbb{U}_d} \left(W_{\{t\}} U_{\{t\}} \right). \tag{19}$$

Remembering the discussion from Sec. 4.2.2, the HiPPI update (19) solves the linearization of problem (7).

4.2.4. Reinterpretation & Criticism

Having this connection uncovered, one can summarize the HiPPI update (11), as first linearizing the objective (8) around the current iterate $U_{\{t\}}$, relaxing the linearized problem by substituting $\sqrt{k}\mathbb{GST}_{nd}$, or rather $\sqrt{k}\mathbb{ST}_{nd}$, for $\mathbb{U}_d$, performing a single orthogonal iteration (12) addressing the relaxed linearization, and then projecting the result back to the set of universe matchings $\mathbb{U}_d$.
Realizing that the projection basically equates to optimizing the linearization of the linearized objective (18) is the foundation of the convergence proof in Sec. 4.3 and offers one last point of view on the HiPPI update (11), see Prop. 1.
Upon reversion, this implies that the originally quartic objective (in $U \in \mathbb{U}_d$) is reduced to a linear one and that problem (7) has been reduced to matrix multiplications and LAPs. This opens up the possibility of criticism. Going from a quartic to a linear objective is an extreme simplification. The benefit of a very simple iteration only based on (efficient) projections and matrix multiplications faces the diverse symmetries of the problem formulation, which become apparent in the, through summation emerging, prefactors in Eq. (9) - the consequence are many redundant operations. Basing the algorithm on the orthogonal iteration but not necessarily initializing it with orthonormal columns removes, strictly speaking, its theoretical foundation, which is (luckily) often not needed, e. g. for convergence, if one adapts the point of view uncovered in Eq. (23) in the proof of Prop. 1.

4.3. Convergence

Momentarily accepting the monotonicity of the sequence $\left(f\left(U_{\{t\}}\right)\right)_{t\in\mathbb{N}}$, shown in Prop. 1, convergence after finite iterations follows from $\mathbb{U}_d$ being finite and $f(U), U \in \mathbb{U}_d$ being bounded above, [1, Corollary 5].

The monotonicity proof in Prop. 1 builds upon the fact that linearizations g_{x_0} of a convex function g are minorizations, $g_{x_0} \le g$, twice. The ideas are the same as in [1, Prop. 3], but stripped down to the essentials, such that generalization becomes easier and the operating principle of the algorithm clearer. Detailed calculations can again be found in Sec. A.4.

Proposition 1. $\forall t \in \mathbb{N}: \quad f\left(U_{\{t\}}\right) \le f\left(U_{\{t+1\}}\right)$

Proof. First of all, we remember the requirement for the weights to be symm. and pos. semidefinite, $W \in \mathbb{S}_n^+$, which in light of Eq. (18) implies $W_{\{t\}} \in \mathbb{S}_n^+$, because, using the decomposition $W = LL^T$, implied by $W \in \mathbb{S}_n^+$, one obtains,

$$W_{\{t\}} = W U_{\{t\}} U_{\{t\}}^T W = \left(LL^T U_{\{t\}}\right)\left(U_{\{t\}}^T LL^T\right) =: \tilde{L}\tilde{L}^T \,. \tag{20}$$

Now, consider the objective, $f : \mathrm{Sym}_n \to \mathbb{R}, X \mapsto \langle WXW, X\rangle_F$, with convex domain $\mathrm{Sym}_n := \left\{M \in \mathbb{R}^{n\times n} \,\middle|\, M = M^T\right\}$, and its linearization in X, $f_{X_{\{t\}}} : \mathbb{R}^{n\times d} \to \mathbb{R}, U \mapsto \langle U, W_{\{t\}} U\rangle_F$.
By checking the pos. semidefiniteness of both their second derivatives, we realize their convexity,

$$\begin{array}{lllll} \partial^2 f(X)(V,V) & = 2\langle WVW, V\rangle_F & = 2\|L^T V L\|_F^2 & \ge & 0 \\ \partial^2 f_{X_{\{t\}}}(U)(V,V) & = 4\langle V, W_{\{t\}} V\rangle_F & = 4\|\tilde{L}^T V\|_F^2 & \ge & 0\,. \end{array} \tag{21}$$

At last, we linearize $f_{X_{\{t\}}}$ in $U_{\{t\}}$,

$$\left(f_{X_{\{t\}}}\right)_{U_{\{t\}}}(U) \sim \langle U, W_{\{t\}} U_{\{t\}}\rangle_F \,, \tag{22}$$

which recovers the term maximized in the projection of the HiPPI update (11),

$$U_{\{t+1\}} \leftarrow \mathrm{proj}_{\mathbb{U}_d}\left(W_{\{t\}} U_{\{t\}}\right) \overset{\text{Eq. (17)}}{=} \underset{U\in\mathbb{U}_d}{\arg\max}\left\{\langle U, W_{\{t\}} U_{\{t\}}\rangle_F\right\} = \underset{U\in\mathbb{U}_d}{\arg\max}\left\{\left(f_{X_{\{t\}}}\right)_{U_{\{t\}}}(U)\right\}. \tag{23}$$

Realizing that linearizations of convex functions are minorizations and agree with the original function at the base point, i.e. $g_{x_0}(x_0) = g(x_0)$, concludes the proof,

$$f\left(U_{\{t\}}\right) \overset{\text{lin.}}{=} f_{X_{\{t\}}}\left(X_{\{t\}}\right) \overset{\text{lin.}}{=} \left(f_{X_{\{t\}}}\right)_{U_{\{t\}}}\left(U_{\{t\}}\right) \overset{\text{Eq. (23)}}{\le} \left(f_{X_{\{t\}}}\right)_{U_{\{t\}}}\left(U_{\{t+1\}}\right) \overset{f_{X_{\{t\}}}\,\text{cvx}}{\le} f_{X_{\{t\}}}\left(X_{\{t+1\}}\right) \overset{f\,\text{cvx}}{\le} f\left(U_{\{t+1\}}\right). \tag{24}$$

□

Notably in Eq. (23), it becomes clear, that the HiPPI update (11) essentialy boils problem (7) down to a LAP in the universe matchings, which given the, by the cost structure implied, convexity leads to an increase of the objective. Hopes for generalizing this convergence result are therefore slim, but one could still investigate how tightly the cost structure and the fact, that linearizing twice reduces the problem to a LAP, are coupled.

4.4. Generalization

In this spirit, one could try to substitute $W \otimes W \in \mathrm{Sym}_{n^2}$ by more general costs $C \in \mathrm{Sym}_{n^2}$, transforming the objective from Eq. (8),

$$f(X) = \langle WXW, X \rangle_F = \langle \vec{X}, W \otimes W \vec{X} \rangle , \tag{25}$$

into the more general objective,

$$F(X) = \langle \vec{X}, C\vec{X} \rangle . \tag{26}$$

Substituting universe matchings according to Lemma 1 brings forth,

$$F(U) = \langle \overrightarrow{UU^T}, C\, \overrightarrow{UU^T} \rangle = \langle \vec{U},\ (U \otimes I_n)^T C\, (U \otimes I_n)\, \vec{U} \ \rangle , \tag{27}$$

while linearization in the manner of Eq. (18) and Eq. (22) in Prop. 1 are,

$$\begin{aligned} F_{X_{\{t\}}}(X) &\sim \langle \vec{X}_{\{t\}}, C\vec{X} \rangle = \langle \vec{U}_{\{t\}},\ \left(U_{\{t\}} \otimes I_n\right)^T C\, (U \otimes I_n)\, \vec{U} \ \rangle \\ \left(F_{X_{\{t\}}}\right)_{U_{\{t\}}}(U) &\sim \langle \vec{U}, \left[\left(U_{\{t\}} \otimes I_n\right) + \left(I_n \otimes U_{\{t\}}\right) K\right]^T C \left(U_{\{t\}} \otimes I_n\right) \vec{U}_{\{t\}} \rangle , \end{aligned} \tag{28}$$

where $K := K_{nd}$ is the $(nd \times nd)$-dimensional commutation matrix, compactly described in [31], whose central property is $K_{nd} \vec{A} = \overrightarrow{A^T}$ for any matrix $A \in \mathbb{R}^{n \times d}$. With a final bit of notation, namely $\mathrm{mat}(\vec{U}) = U$, this last linearization suggests the update

$$U_{\{t+1\}} \leftarrow \mathrm{proj}_{\mathbb{U}_d} \left(\mathrm{mat} \left(\left[\left(U_{\{t\}} \otimes I_n\right) + \left(I_n \otimes U_{\{t\}}\right) K\right]^T C \left(U_{\{t\}} \otimes I_n\right) \quad \vec{U}_{\{t\}} \right) \right) , \tag{29}$$

which reduces to the HiPPI Update (11) for the choice $C = W \otimes W$ as is expected and sanity checked in Sec. A.5. This is an even more expensive update than the original HiPPI update (11), a yet mathematically unfounded, simplification would be to drop the therm involving $\left(I_n \otimes U_{\{t\}}\right) K$ and instead solve the problem

$$\max_{U \in \mathbb{U}_d} \left\{ \langle \vec{U},\ \left(U_{\{t\}} \otimes I_n\right)^T C \left(U_{\{t\}} \otimes I_n\right) \vec{U} \ \rangle \right\} , \tag{30}$$

Experimental evaluation and further investigation is needed to determine its practicality. Another direction of future work could be simplifying the update in the case of pairwise decomposable costs.

5. Notes on Experiments

HiPPI was compared on three datasets (HiPPI-Dataset [1], WILLOW [32], TOSCA [33]) to a plethora of other methods. These other methods were mostly designed with in part heavily differing contexts in mind, just as HiPPI is tailored towards a certain structure of problems. All datasets are motivated from a computer vision perspective and are either shapes or images. The construction touched upon in Sec. 4.1.1 was conducted for all datasets, where either 2D Euclidean space, with canonical metric, or the shape manifold, with geodesic distances, provided the structure of a metric space. Naturally, modelling and optimization of the problem are intertwined, making an actual evaluation of the optimization method difficult. In accordance, energy or costs of the final solution were never used as a metric of comparison, but instead

e. g. precision and recall, which rest upon veiled ground truth. Detailed setups and results should be read in [1]. The following just sheds light on possible flaws of the comparisons.
Experiments on the HiPPI-Dataset are conducted between HiPPI and methods not incorporating any geometric relations, let alone being able to handle costs relating pairs of matchings - quite contrary to HiPPI. Although this flaw is mentioned by Bernard et al. and excused by inefficiencies of the methods that do incorporate geometric relations, the comparison and conclusions drawn from this experiment are not less than questionable.
Besides the overall criticism regarding the experiments, data obtained via WILLOW [32] allows a somewhat fair comparison. Solely the simplicity of the dataset and corresponding problems could be remarked.
The comparison on TOSCA [33] seems to be misguided, as the only method included in the comparison, [34] by Cosmo et al., approaches a vastly more difficult problem - outlier shapes. Their two-stage filtering approach, aimed at these problems, immediately provides a seemingly straightforward explanation for the in [1] observed (and condemned) low number of matches. Whilst this is also mentioned by Bernard et al., the comparison should be taken with a grain of salt - HiPPI would fail in the scenarios [34] was designed to handle.
Arguing for the benefit of Bernard et al.'s experimental evaluation is the fact, that, especially in the literary landscape at the time, methods targeting this problem (from the computer vision perspective) mix, as mentioned, modelling and optimization. Each modelling stage uses slightly different assumptions outside of which applicability is often lost - this makes a 'fair' comparison almost impossible. Both stages need to be split. Then again, a lot of details, that could help to reduce this gap to a 'fair' comparison, were swept under the rug. Just to mention a few: Examples of the HiPPI dataset as well as ground truth are not provided. Initialisation, e. g. of permutations synchronization methods, if at all mentioned, was done with unspecified linear weights, while HiPPI performs local search using quadratic weights. Just as grave, it was left unclear, whether weights used by HiPPI were reused for methods, that, for a change, are able to handle quadratic weights.

6. Conclusion

After a well-founded mathematical analysis, HiPPI was unveiled as simply optimizing a Multi-LAP obtained by twofold linearization of the Multi-QAP objective . This perspective paved the way to a generalization of the HiPPI update, which, more than anything, stressed the essentiality of the cost structure used by Bernard et al. in [1], as it entails convexity - with it the basis for the method's convergence guarantee, as well as efficiency by relying on lower dimensional matrices. Nonetheless, the generalization opened up exciting new avenues for future work, of which experimental evaluation poses as the most crucial one.

7. References

[1] Florian Bernard et al. "Hippi: Higher-order projected power iterations for scalable multi-matching". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, pp. 10284–10293.

[2] Rainer Burkard, Mauro Dell'Amico, and Silvano Martello. *Assignment problems: revised reprint*. SIAM, 2012.

[3] James Munkres. "Algorithms for the assignment and transportation problems". In: *Journal of the society for industrial and applied mathematics* 5.1 (1957), pp. 32–38.

[4] Jack Edmonds and Richard M Karp. "Theoretical improvements in algorithmic efficiency for network flow problems". In: *Journal of the ACM (JACM)* 19.2 (1972), pp. 248–264.

[5] Dimitri P Bertsekas. "A distributed algorithm for the assignment problem". In: *Lab. for Information and Decision Systems Working Paper, MIT* (1979).

[6] BL Schwartz. "A computational analysis of the auction algorithm". In: *European journal of operational research* 74.1 (1994), pp. 161–169.

[7] Florian Bernard et al. "Fast correspondences for statistical shape models of brain structures". In: *Medical Imaging 2016: Image Processing*. Vol. 9784. SPIE. 2016, pp. 197–204.

[8] Eugene L Lawler. "The quadratic assignment problem". In: *Management science* 9.4 (1963), pp. 586–599.

[9] Olivier Duchenne et al. "A tensor-based algorithm for high-order graph matching". In: *IEEE transactions on pattern analysis and machine intelligence* 33.12 (2011), pp. 2383–2395.

[10] Xinchu Shi et al. "Tensor power iteration for multi-graph matching". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 5062–5070.

[11] Tjalling C Koopmans and Martin Beckmann. "Assignment problems and the location of economic activities". In: *Econometrica: journal of the Econometric Society* (1957), pp. 53–76.

[12] Panos M Pardalos, Ding-Zhu Du, and Ronald L Graham. *Handbook of combinatorial optimization*. Springer, 2013.

[13] Stefan Haller et al. "A comparative study of graph matching algorithms in computer vision". In: *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXIII*. Springer. 2022, pp. 636–653.

[14] Paul Swoboda et al. "A convex relaxation for multi-graph matching". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 11156–11165.

[15] Olivia Reynen et al. "Large-scale multi-dimensional assignment: Problem formulations and gpu accelerated solutions". In: *2019 22th International Conference on Information Fusion (FUSION)*. IEEE. 2019, pp. 1–8.

[16] Yves Crama and Frits CR Spieksma. "Approximation algorithms for three-dimensional assignment problems with triangle inequalities". In: *European Journal of Operational Research* 60.3 (1992), pp. 273–279.

[17] Deepti Pachauri, Risi Kondor, and Vikas Singh. “Solving the multi-way matching problem by permutation synchronization”. In: *Advances in neural information processing systems* 26 (2013).

[18] Xiaowei Zhou, Menglong Zhu, and Kostas Daniilidis. “Multi-image matching via fast alternating minimization”. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 4032–4040.

[19] Eleonora Maset, Federica Arrigoni, and Andrea Fusiello. “Practical and efficient multi-view matching”. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2017, pp. 4568–4576.

[20] Roberto Tron et al. “Fast multi-image matching via density-based clustering”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 4057–4066.

[21] Junchi Yan et al. “Multi-graph matching via affinity optimization with graduated consistency regularization”. In: *IEEE transactions on pattern analysis and machine intelligence* 38.6 (2015), pp. 1228–1242.

[22] Itay Kezurer et al. “Tight relaxation of quadratic matching”. In: *Computer graphics forum*. Vol. 34. 5. Wiley Online Library. 2015, pp. 115–128.

[23] Florian Bernard, Christian Theobalt, and Michael Moeller. “Ds*: Tighter lifting-free convex relaxations for quadratic matching problems”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 4310–4319.

[24] Junchi Yan et al. “A matrix decomposition perspective to multiple graph matching”. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 199–207.

[25] Qianqian Wang, Xiaowei Zhou, and Kostas Daniilidis. “Multi-image semantic matching by mining consistent features”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 685–694.

[26] Gene H. Golub and Charles F. Van Loan. “Matrix computations.” In: (1996).

[27] Florian Bernard, Daniel Cremers, and Johan Thunberg. “Sparse quadratic optimisation over the Stiefel manifold with application to permutation synchronisation”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 25256–25266.

[28] Nicholas J Higham. *Functions of matrices: theory and computation*. SIAM, 2008.

[29] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. “Optimization algorithms on matrix manifolds”. In: *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2009.

[30] Yuxin Chen and Emmanuel J Candès. “The projected power method: An efficient algorithm for joint alignment from pairwise differences”. In: *Communications on Pure and Applied Mathematics* 71.8 (2018), pp. 1648–1714.

[31] Jan R Magnus and Heinz Neudecker. “The commutation matrix: some properties and applications”. In: *The annals of statistics* 7.2 (1979), pp. 381–394.

[32] Minsu Cho, Karteek Alahari, and Jean Ponce. “Learning graphs to match”. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2013, pp. 25–32.

[33] Rui Xiang, Rongjie Lai, and Hongkai Zhao. “A dual iterative refinement method for non-rigid shape matching”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 15930–15939.

[34] Luca Cosmo et al. "Consistent partial matching of shape collections via sparse modeling". In: *Computer Graphics Forum*. Vol. 36. 1. Wiley Online Library. 2017, pp. 209–221.

(a) Existing universe matching for $k = 2$.

(b) Construction for $k = 2 \Rightarrow k = 3$.

Figure 1: Visualization of Lemma A.1. The elements of the $k = 3$ sets $\mathcal{V}^{[p]}$ are represented by circles underneath the corresponding set. The assumed matching $X \in \mathbb{X}$ is indicated via lines, where a line between elements means $X^{[p,q]}_{i_p i_q} = 1$. One should take note of the cycle-consistency. The universe is denoted in the most left column of each picture, whereas matchings to the universe are visualized by the color. Fig. 1a shows the induction hypothesis and Fig. 1b the final construction.

A. Appendix

A.1. Universe Characterization

The idea behind the construction used in the proof of Lemma A.1 is visualized in Fig. 1. Intuitively one assigns to each element participating in a maximal cycle the same universe point. In this context, a maximal cycle is the set of elements obtained by starting with the singleton of an arbitrary element and adding each element matched to an already added element.

Lemma A.1. *(Universe Characterization of Cycle-Consistency)*
Assume $X \in \left[\mathbb{P}^{[p,q]}\right]_{p,q \in [\![k]\!]^2}$ *, then* $X \in \mathbb{X} \quad \Leftrightarrow \quad \exists d \in \mathbb{N} \colon \exists U \in \mathbb{U}_d \colon \forall p, q \in [\![k]\!] \colon X^{[p,q]} = U^{[p]} \left(U^{[q]}\right)^T,$
where $\mathbb{U}_d := \left\{ \left[U^{[p]}\right]_{p \in [\![k]\!]} \in \left[\mathbb{P}_{n_p, d}\right]_{p \in [\![k]\!]} \middle| \forall p \in [\![k]\!] : U^{[p]} \vec{1}_d = \vec{1}_{n_p} \right\}$ *denotes the set of universe matchings for a universe of size* d*. Rephrased, it holds that* $X = UU^T$.

Proof. For "$\Leftarrow$", see [1, Lemma 2] and for "$\Rightarrow$", inductive proof per construction in $k \in \mathbb{N}$.

$\underline{k = 2:}$
Denote the unassigned elements of $\mathcal{V}^{[2]}$ with $\mathcal{W}^{[2]} := \left\{ i_2 \in \mathcal{V}^{[2]} \middle| \left[\left(X^{[1,2]}\right)^T \vec{1}_{n_1} \right]_{i_2} = 0 \right\}$.
Set the universe size to $d := \underbrace{|\mathcal{V}^{[1]}|}_{=:d^{[1]}} + \underbrace{|\mathcal{W}^{[2]}|}_{=:d^{[2]}}$ and the universe matching of $\mathcal{V}^{[1]}$ to $U^{[1]} := \left[I_{n_1} \quad 0_{n_1 \times d^{[2]}} \right]$.

We note the matching relations following from $X^{[1,2]} \in \{0,1\}^{n_1 \times n_2}$ and the definition of $\mathscr{W}^{[2]}$

- $\forall i_1 \in \mathscr{V}^{[1]} \exists! (i_d)_{i_1} \in [\![d^{[1]}]\!] : U^{[1]}_{i_1 (i_d)_{i_1}} = 1$
- $\forall i_2 \in \mathscr{V}^{[2]} \setminus \mathscr{W}^{[2]} \exists! (i_1)_{i_2} \in \mathscr{V}^{[1]} : X^{[1,2]}_{(i_1)_{i_2} i_2} = 1,$

and construct the second, thereby well defined, universe matching

- $\left[U^{[2]}_{i_2 i_d}\right]_{i_2, i_d \in \mathscr{W}^{[2]} \times [\![d]\!]} := \begin{bmatrix} 0_{d^{[2]} \times d^{[1]}} & I_{d^{[2]}} \end{bmatrix}$
- $\left[U^{[2]}_{i_2 i_d}\right]_{i_2, i_d \in \mathscr{V}^{[2]} \setminus \mathscr{W}^{[2]} \times [\![d]\!]} := \left[\begin{cases} 1; & i_d = (i_d)_{(i_1)_{i_2}} \\ 0; & \text{else} \end{cases}\right]_{i_2, i_d \in \mathscr{V}^{[2]} \setminus \mathscr{W}^{[2]} \times [\![d]\!]}.$

With this construction it holds that $U := \begin{bmatrix} U^{[1]} \\ U^{[2]} \end{bmatrix} \in \mathbb{U}_d$ and

$$\left[U^{[1]} \left(U^{[2]}\right)^T\right]_{i_1 i_2} = \sum_{i_d \in [\![d]\!]} U^{[1]}_{i_1 i_d} U^{[2]}_{i_2 i_d} = \begin{cases} 1; & i_d = (i_d)_{i_1} \wedge i_d = (i_d)_{(i_1)_{i_2}} \Leftrightarrow i_1 = (i_1)_{i_2} \Leftrightarrow X^{[1,2]}_{i_1 i_2} = 1 \\ 0; & \text{else} \end{cases} = X^{[1,2]}_{i_1 i_2}.$$

Hence, with the assumed cycle-consistency of X, $X = UU^T$.

<u>$k-1 \Rightarrow k$:</u> Assume the existence of universe matchings $\tilde{U}^{[p]}$ for $p \in [\![k-1]\!]$ to a universe of size d_{k-1}.

Denote the unassigned elements of $\mathscr{V}^{[k]}$ with $\mathscr{W}^{[k]} := \left\{ i_k \in \mathscr{V}^{[k]} \,\middle|\, \forall p \in [\![k-1]\!] : \left[\left(X^{[p,k]}\right)^T \vec{1}_{n_p}\right]_{i_k} = 0 \right\}$.
Now noting the matching relations

- $\forall p \in [\![k-1]\!] \forall i_p \in \mathscr{V}^{[p]} \exists! (i_d)_{i_p} \in [\![d_{k-1}]\!] : \tilde{U}^{[p]}_{i_p (i_d)_{i_p}} = 1$
- $\forall i_k \in \mathscr{V}^{[k]} \setminus \mathscr{W}^{[k]} : \emptyset \neq \mathscr{I}_{i_k} := \left\{ p \in [\![k-1]\!] \,\middle|\, \left[\left(X^{[p,k]}\right)^T \vec{1}_{n_p}\right]_{i_k} = 1 \right\} \rightsquigarrow \forall p \in \mathscr{I}_{i_k} : \exists! \left(i_p\right)_{i_k} : X^{[p,k]}_{(i_p)_{i_k} i_k} = 1,$

transitivity of cycle-consistency demands that $\forall i_k \in \mathscr{V}^{[k]} \setminus \mathscr{W}^{[k]} \forall p, q \in \mathscr{I}_{i_k} : (i_d)_{(i_p)_{i_k}} = (i_d)_{(i_q)_{i_k}} =: (i_d)_{i_k}$.
This can be shown by contradiction.
Assuming $\exists p \neq q \in \mathscr{I}_{i_k} : (i_d)_{(i_p)_{i_k}} \neq (i_d)_{(i_q)_{i_k}}$ $(*)$ and realizing $X^{[p,q]}_{(i_p)_{i_k}, (i_q)_{i_k}} = \sum_{i_d \in [\![d_{k-1}]\!]} \tilde{U}^{[p]}_{(i_p)_{i_k} i_d} \tilde{U}^{[q]}_{(i_q)_{i_k} i_d} \overset{(*)}{=} 0$,
as well as $\left[X^{[p,k]} X^{[k,q]}\right]_{(i_p)_{i_k}, (i_q)_{i_k}} = \sum_{j_k \in \mathscr{V}^{[k]}} X^{[p,k]}_{(i_p)_{i_k}, j_k} X^{[q,k]}_{(i_q)_{i_k}, j_k} = \underbrace{X^{[p,k]}_{(i_p)_{i_k}, i_k}}_{=1} \underbrace{X^{[q,k]}_{(i_q)_{i_k}, i_k}}_{=1} = 1$
reveals the contradiction to $X^{[p,k]} X^{[k,q]} \leq X^{[p,q]}$.
Finally, the universe matchings can be constructed. Set the universe size to $d := d_{k-1} + \underbrace{|\mathscr{W}^{[k]}|}_{=: d^{[k]}}$ and define

- $\forall p \in [\![k-1]\!] : U^{[p]} := \begin{bmatrix} \tilde{U}^{[p]} & 0_{n_p \times d^{[k]}} \end{bmatrix}$

- $\left[U^{[k]}_{i_k i_d}\right]_{i_k, i_d \in \mathcal{W}^{[k]} \times [\![d]\!]} := \left[0_{d^{[k]} \times d_{k-1}} \quad I_{d^{[k]}}\right]$

- $\left[U^{[k]}_{i_k i_d}\right]_{i_k, i_d \in \mathcal{V}^{[k]} \setminus \mathcal{W}^{[k]} \times [\![d]\!]} := \left[\begin{Bmatrix} 1; & i_d = (i_d)_{i_k} \\ 0; & \text{else} \end{Bmatrix}\right]_{i_k, i_d \in \mathcal{V}^{[k]} \setminus \mathcal{W}^{[k]} \times [\![d]\!]}$

Then,

- $U := \left[U^{[p]}\right]_{p \in [\![k]\!]} \in \mathbb{U}_d$

- $\forall p, q \in [\![k-1]\!] : U^{[p]} \left(U^{[q]}\right)^T = \tilde{U}^{[p]} \left(\tilde{U}^{[q]}\right)^T = X^{[p,q]}$

- $\forall p \in [\![k-1]\!] : \left[U^{[p]} \left(U^{[k]}\right)^T\right]_{i_p i_k} = \sum_{i_d \in [\![d]\!]} U^{[p]}_{i_p i_d} U^{[k]}_{i_k i_d} = \begin{Bmatrix} 1; & i_p = (i_p)_{i_k} \Leftrightarrow X^{[p,k]}_{i_p i_k} = 1 \\ 0; & \text{else} \end{Bmatrix} = X^{[p,k]}_{i_p i_k}$,

which means $X = UU^T$ and concludes the proof. □

A.2. Details: Problem Formulation

In the following the substitution of universe matchings from Eq. (8) is done with more rigor.

$$
\begin{aligned}
f(U) &= \langle U^T W U, U^T W U \rangle_F \\
&= \langle \left(UU^T\right) W, W \left(UU^T\right) \rangle_F \\
&\overset{X=UU^T}{=} \langle XW, WX \rangle_F \\
&\overset{W=W^T}{=} \langle WXW, X \rangle_F \\
&\overset{X=X^T}{=} \operatorname{tr}(WXWX) \\
&= \sum_{p \in [\![k]\!]} \operatorname{tr}\left([WXWX]^{[p,p]}\right) \\
&= \sum_{p \in [\![k]\!]} \operatorname{tr}\left(\sum_{(q,r,s) \in [\![k]\!]^3} W^{[p,q]} X^{[q,r]} W^{[r,s]} X^{[s,p]}\right) \\
&= \sum_{(p,q,r,s) \in [\![k]\!]^4} \operatorname{tr}\left(W^{[p,q]} X^{[q,r]} W^{[r,s]} X^{[s,p]}\right) \\
&= \sum_{(p,q,r,s) \in [\![k]\!]^4} \langle X^{[s,p]}, W^{[s,r]} X^{[r,q]} W^{[q,p]} \rangle_F \\
&= \sum_{(p,q,r,s) \in [\![k]\!]^4} \langle \overrightarrow{X^{[s,p]}}, \overrightarrow{W^{[s,r]} X^{[r,q]} W^{[q,p]}} \rangle \\
&= \sum_{(p,q,r,s) \in [\![k]\!]^4} \langle \overrightarrow{X^{[s,p]}}, \left(W^{[q,p]}\right)^T \otimes W^{[s,r]} \overrightarrow{X^{[r,q]}} \rangle.
\end{aligned}
\tag{31}
$$

The special case of a Multi-QAP with decomposable costs in Koopmanns-Beckmann's form proclaimed in Eq. (10) in the unweighted scenario, is recovered as follows. Assume $S = I_n$, which implies $W = A$, and recall, that A is of block-diagonal form $A = \operatorname{diag}\left(A^{[p]}, p \in [\![k]\!]\right) \in \mathbb{S}_n^+$.

$$
\begin{aligned}
f(X) &= \langle WXW, X\rangle_F \\
&\overset{W=A}{=} \langle AXA, X\rangle_F \\
&\overset{\text{see Eq. (31)}}{=} \sum_{(p,q,r,s)\in[\![k]\!]^4} \operatorname{tr}\left(A^{[p,q]} X^{[q,r]} A^{[r,s]} X^{[s,p]}\right) \\
&\overset{p\neq q \Rightarrow A^{[p,q]}=0_{n_p\times n_q}}{=} \sum_{(p,s)\in[\![k]\!]^2} \operatorname{tr}\left(A^{[p]} X^{[p,s]} A^{[s]} X^{[s,p]}\right) \\
&\overset{\text{see Eq. (31)}}{=} \sum_{(p,s)\in[\![k]\!]^2} \langle \overrightarrow{X^{[p,s]}}, \left(A^{[s]}\right)^T \otimes A^{[p]} \overrightarrow{X^{[p,s]}}\rangle \\
&\overset{A^{[p]}=\left(A^{[p]}\right)^T}{=} \sum_{(p,s)\in[\![k]\!]^2} \langle \overrightarrow{X^{[p,s]}}, A^{[s]} \otimes A^{[p]} \overrightarrow{X^{[p,s]}}\rangle.
\end{aligned}
\tag{32}
$$

A.3. Details: Relaxations

Proofs for Lemmas 2 - 4.

Lemma A.2. $\max\limits_{U\in\sqrt{k}\mathbb{GST}_{nd}} \{\langle U, WU\rangle_F\}$ *is a relaxation of* $\max\limits_{U\in\mathbb{U}_d} \{\langle U, WU\rangle_F\}$.

Proof. Choose $U \in \mathbb{U}_d$, then $U^{[p]} \in \mathbb{P}_{n_p d} \Rightarrow \left(U^{[p]}\right)^T U^{[p]} \le I_d$.
Now, $U^T U = \sum\limits_{p\in[\![k]\!]} \left(U^{[p]}\right)^T U^{[p]} \le \sum\limits_{p\in[\![k]\!]} I_d = kI_d \quad \Rightarrow \quad \mathbb{U}_d \subseteq \sqrt{k}\mathbb{GST}_{nd}$.

□

Lemma A.3. $\operatorname*{arg\,max}\limits_{U\in\sqrt{k}\mathbb{GST}_{nd}} \{\langle U, WU\rangle_F\} \overset{1}{=} \sqrt{k} \operatorname*{arg\,max}\limits_{U\in\mathbb{GST}_{nd}} \{\langle U, WU\rangle_F\} \overset{2}{=} \sqrt{k} \operatorname*{arg\,max}\limits_{U\in\mathbb{ST}_{nd}} \{\langle U, WU\rangle_F\}$ *, where the second equality requires* $W \in \mathbb{S}_n^+$ *with eigenvalues* $\lambda_1, \ldots, \lambda_d > \lambda_{d+1}, \ldots, \lambda_n$.

Proof.

To 1,

$$\text{LHS} = \sqrt{k} \operatorname*{arg\,max}_{U\in\mathbb{GST}_{nd}} \left\{\langle(\sqrt{k}U), W(\sqrt{k}U)\rangle_F\right\} = \sqrt{k} \operatorname*{arg\,max}_{U\in\mathbb{GST}_{nd}} \{k\langle U, WU\rangle_F\} = \text{RHS} .$$

To 2, notice,

$$\langle U, WU\rangle_F = \sum_{i_d\in[\![d]\!]} \langle U_{\cdot i_d}, WU_{\cdot i_d}\rangle_F.$$

It suffices to show $\forall \widetilde{U} \in \mathbb{GST}_{nd} \setminus \mathbb{ST}_{nd} \exists U \in \mathbb{ST}_{nd} : f(U) > f(\widetilde{U})$.
Choose $\widetilde{U} \in \mathbb{GST}_{nd} \setminus \mathbb{ST}_{nd}$, and define the index sets $\mathscr{I} := \left\{ i \in [\![d]\!] : \|\widetilde{U}_{\cdot i}\|^2 < 1 \right\}$ and $\mathscr{N} := \left\{ i \in \mathscr{I} : \widetilde{U}_{\cdot i} = 0 \right\}$.

$\underline{\mathscr{N} = \emptyset}$: Define $U_{\cdot i} := \left\{ \begin{array}{ll} \frac{\widetilde{U}_{\cdot i}}{\|\widetilde{U}_{\cdot i}\|}; & i \in \mathscr{I} \\ \widetilde{U}_{\cdot i}; & i \notin \mathscr{I} \end{array} \right\}$. Then, $U \in \mathbb{ST}_{nd}$ and $f(U) = \underbrace{\left(\prod_{i \in \mathscr{I}} \|\widetilde{U}_{\cdot i}\|^{-2} \right)}_{>1} f(\widetilde{U}) > f(\widetilde{U})$.

$\underline{\mathscr{N} \neq \emptyset}$: We will show that we can find $|\mathscr{N}|$ objective increasing, non-zero vectors, that remain orthogonal to the non-zero columns and each other. Denote the eigenpairs of W with $(\lambda_i, w_i)_{i \in [\![n]\!]}$. Define the set of non-zero columns $\mathscr{U} := \left\{ \widetilde{U}_{\cdot i} | i \in [\![d]\!] \setminus \mathscr{N} \right\} \subseteq \mathbb{R}^n$, and leading eigenvectors $\mathscr{W} := \left\{ w_i | i \in [\![d]\!] \right\} \subseteq \mathbb{R}^n$. Mutual orthogonality of vectors in $\mathscr{U}$, implied by $\widetilde{U} \in \mathbb{GST}_{nd}$, implies linear independence, which in turn implies $\dim\left(\operatorname{span}(\mathscr{U})\right) = d - |\mathscr{N}|$. Now,

$$\begin{aligned} \dim\left(\operatorname{span}(\mathscr{U})^{\perp} \cap \operatorname{span}(\mathscr{W})\right) &= \underbrace{\dim\left(\operatorname{span}(\mathscr{U})^{\perp}\right)}_{=n-d+|\mathscr{N}|} + \underbrace{\dim\left(\operatorname{span}(\mathscr{W})\right)}_{=d} - \underbrace{\dim\left(\operatorname{span}(\mathscr{U})^{\perp} + \operatorname{span}(\mathscr{W})\right)}_{\leq n} \\ &\geq |\mathscr{N}| . \end{aligned}$$

Choose any orthonormal basis $\{v_i\}_{i \in [\![|\mathscr{N}|]\!]} \in \operatorname{span}(\mathscr{U})^{\perp} \cap \operatorname{span}(\mathscr{W})$ and define $U_{\cdot i} := \left\{ \begin{array}{ll} v_i; & i \in \mathscr{N} \\ \widetilde{U}_{\cdot i}; & i \notin \mathscr{N} \end{array} \right\}$. Then,

$$f(U) = \sum_{i \in [\![d]\!] \setminus \mathscr{N}} \langle \widetilde{U}_{\cdot i}, W \widetilde{U}_{\cdot i} \rangle_F + \sum_{i \in [\![|\mathscr{N}|]\!]} \underbrace{\langle v_i, W v_i \rangle_F}_{\geq \lambda_d > 0} > f(\widetilde{U}).$$

To ensure $U \in \mathbb{ST}_{nd}$ one possibly has to proceed analogously to the case $\underline{\mathscr{N} = \emptyset}$. □

Lemma A.4. $\operatorname{proj}_{\mathbb{U}_d}\left(\sqrt{k} V\right) = \operatorname{proj}_{\mathbb{U}_d}(V)$

Proof.

$$\begin{aligned} \text{LHS} &= \operatorname{argmin}_{U \in \mathbb{U}_d} \left\{ \|\sqrt{k} V - U\|_F^2 = \langle \sqrt{k} V - U, \sqrt{k} V - U \rangle_F = k\|V\|_F^2 + \|U\|_F^2 - 2\sqrt{k} \langle V, U \rangle_F \right\} \\ &\overset{\|U\|_F^2 = n}{=} \operatorname{argmax}_{U \in \mathbb{U}_d} \left\{ \sqrt{k} \langle V, U \rangle_F \right\} = \operatorname{argmax}_{U \in \mathbb{U}_d} \{ \langle V, U \rangle_F \} \end{aligned}$$

Analogously, RHS $= \operatorname{argmax}_{U \in \mathbb{U}_d} \{ \langle V, U \rangle_F \}$. □

A.4. Details: Derivatives & Linearizations

In the following, calculations, omitted for brevity throughout the report, are carried out in detail.
Recalling the reformulated objective function (8),

$$f : \operatorname{Sym}_n \to \mathbb{R},\ X \mapsto \langle W X W, X \rangle_F \tag{33}$$

with weights $W \in \mathbb{S}_n^+$, allows for the calculation of the first,

$$\partial f(X_0)(V) = \lim_{\alpha \to 0} \left\{ \frac{1}{\alpha} \left(f(X_0 + \alpha V) - f(X_0) \right) \right\} = 2\langle W X_0 W, V \rangle_F \,, \tag{34}$$

and second derivative in $X_0 \in \mathrm{Sym}_n$, along $V \in \mathrm{Sym}_n$,

$$\begin{aligned} \partial^2 f(X_0)(V) &= \lim_{\alpha \to 0} \left\{ \tfrac{1}{\alpha} \left(\partial f(X_0 + \alpha V) - \partial f(X_0) \right) \right\} \\ &= \lim_{\alpha \to 0} \left\{ \tfrac{2}{\alpha} \left(\langle W \left(X_0 + \alpha V \right) W, \cdot \rangle_F - \langle W X_0 W, \cdot \rangle_F \right) \right\} \\ &= 2\langle W V W, \cdot \rangle_F \\ \partial^2 f(X_0)(V, V) &= 2\langle W V W, V \rangle_F \,, \end{aligned} \tag{35}$$

where $\cdot$ indicates that the objective is a function awaiting an argument.
This gives rise to the linearization $f_{X_{\{t\}}}$ of f in $X_{\{t\}} \in \mathrm{Sym}_n$,

$$f_{X_{\{t\}}}(X) = f(X_{\{t\}}) + \partial f(X_{\{t\}})(X - X_{\{t\}}) = f(X_{\{t\}}) + 2\langle W X_{\{t\}} W, X - X_{\{t\}} \rangle_F := h(X_{\{t\}}) + 2\langle W X_{\{t\}} W, X \rangle_F \,, \tag{36}$$

with the function h summarizing all terms constant w.r.t. X. Using Lemma 1 to reformulate this in terms of universe matchings, while recalling $W_{\{t\}} := W X_{\{t\}} W = W U_{\{t\}} U_{\{t\}}^T W \in \mathrm{Sym}_n$from Eq. (18),

$$f_{X_{\{t\}}} : \mathbb{R}^{n \times d} \to \mathbb{R}, U \mapsto h(X_{\{t\}}) + 2\langle U, W_{\{t\}} U \rangle_F \,, \tag{37}$$

allows to calculate the corresponding derivatives in $U_0 \in \mathbb{R}^{n \times d}$, along $V \in \mathbb{R}^{n \times d}$,

$$\begin{aligned} \partial f_{X_{\{t\}}}(U_0)(V) &= 4\langle U_0, W_{\{t\}} V \rangle_F \\ \partial^2 f_{X_{\{t\}}}(U_0)(V, V) &= 4\langle V, W_{\{t\}} V \rangle_F \end{aligned} \quad , \tag{38}$$

giving rise to the last linearization,

$$\left(f_{X_{\{t\}}} \right)_{U_{\{t\}}}(U) = f_{X_{\{t\}}}\left(U_{\{t\}} \right) + \partial f_{X_{\{t\}}}(U_{\{t\}})(U - U_{\{t\}}) = f_{X_{\{t\}}}\left(U_{\{t\}} \right) + 4\langle U_{\{t\}}, W_{\{t\}} U \rangle_F \,. \tag{39}$$

Now turning to the generalized objective from Eq. (26),

$$F : \mathrm{Sym}_n \to \mathbb{R},\, X \mapsto \langle \vec{X}, C \vec{X} \rangle \,, \tag{40}$$

with generalized weights $C \in \mathrm{Sym}_{n^2}$. Tracing the exact same calculations, its derivative in $X_0 \in \mathrm{Sym}_n$ along $V \in \mathrm{Sym}_n$ reads,

$$\partial F(X_0)(V) = \lim_{\alpha \to 0} \left\{ \frac{1}{\alpha} \left(\langle \overrightarrow{X_0 + \alpha V}, C \overrightarrow{X_0 + \alpha V} \rangle - \langle \vec{X}_0, C \vec{X}_0 \rangle \right) \right\} \overset{\overrightarrow{A+B} = \vec{A} + \vec{B}}{=} 2\langle \vec{X}_0, C \vec{V} \rangle \,, \tag{41}$$

resulting in the linearization,

$$F_{X_{\{t\}}}(X) = F(X_{\{t\}}) + 2\langle \vec{X}_{\{t\}}, C \left(\vec{X} - \vec{X}_{\{t\}} \right) \rangle := H(X_{\{t\}}) + 2\langle \vec{X}_{\{t\}}, C \vec{X} \rangle \,, \tag{42}$$

this time, with the function H summarizing all terms constant w.r.t. X. Reformulation via Lemma 1 as in Eq. (27) reads,

$$F_{X_{\{t\}}} : \mathbb{R}^{n \times d} \to \mathbb{R}, U \mapsto H(X_{\{t\}}) + 2 \, \langle \vec{U}_{\{t\}}, \left(U_{\{t\}} \otimes I_n \right)^T C \left(U \otimes I_n \right) \vec{U} \rangle \,, \tag{43}$$

while its derivative in $U_0 \in \mathbb{R}^{n\times d}$ along $V \in \mathbb{R}^{n\times d}$ reads, using the bilinearity of the kronecker product,

$$\partial F_{X_{\{t\}}}(U_0)(V) = 2\langle \vec{U}_{\{t\}}, \left(U_{\{t\}} \otimes I_n\right)^T C\left[(V \otimes I_n)\vec{U}_0 + (U_0 \otimes I_n)\vec{V}\right]\rangle . \tag{44}$$

The conclusion is its linearization,

$$\left(F_{X_{\{t\}}}\right)_{U_{\{t\}}}(U) = F_{X_{\{t\}}}(U_{\{t\}}) + 2\left(\underbrace{\langle \vec{U}_{\{t\}}, \left(U_{\{t\}} \otimes I_n\right)^T C\,(U \otimes I_n)\,\vec{U}_{\{t\}}\rangle}_{①} + \underbrace{\langle \vec{U}_{\{t\}}, \left(U_{\{t\}} \otimes I_n\right)^T C\left(U_{\{t\}} \otimes I_n\right)\vec{U}\rangle}_{②}\right). \tag{45}$$

While term ② can be adressed by a LAP analogously to the term in Eq. (22) emerging in the original HiPPI objective, term ① needs to be rearranged. This lastly requires the $(nd \times nd)$-dimensional commutation matrix K_{nd}, see [31]. To avoid confusion, the indices are dropped wherever possible, $K := K_{n,d}$.

$$\begin{aligned} ① &= \langle \vec{U}_{\{t\}}, \left(U_{\{t\}} \otimes I_n\right)^T C\,(U \otimes I_n)\,\vec{U}_{\{t\}}\rangle && ;\text{ with } \left((U \otimes I_n)\,\vec{U}_{\{t\}} = \overrightarrow{U_{\{t\}}U^T} = \left(I_n \otimes U_{\{t\}}\right)\overrightarrow{U^T}\right) \\ &= \langle \vec{U}_{\{t\}}, \left(U_{\{t\}} \otimes I_n\right)^T C\left(I_n \otimes U_{\{t\}}\right)\overrightarrow{U^T}\rangle && ;\text{ with } \left(K\vec{U} = \overrightarrow{U^T}\right) \\ &= \langle \vec{U}_{\{t\}}, \left(U_{\{t\}} \otimes I_n\right)^T C\left(I_n \otimes U_{\{t\}}\right)K\vec{U}\rangle \end{aligned} \tag{46}$$

The second linearization of the generalized objective then takes the form,

$$\left(F_{X_{\{t\}}}\right)_{U_{\{t\}}}(U) = F_{X_{\{t\}}}(U_{\{t\}}) + 2 \quad \langle \vec{U}, \left[\left(U_{\{t\}} \otimes I_n\right) + \left(I_n \otimes U_{\{t\}}\right)K\right]^T C\left(U_{\{t\}} \otimes I_n\right)\vec{U}_{\{t\}}\rangle . \tag{47}$$

A.5. Reduction of the Generalized to the HiPPI Update

The goal is to show that the generalized update (29) reduces to the HiPPI update (11) (in form (19)) under the choice of $C = W \otimes W$ for $W \in \mathrm{Sym}_n$. The calculation uses the fact that the commutation matrix K is a permutation matrix and as such orthogonal, $K^T K = I_{nd}$ [31, Thm. 3.1 (iii)].

$$\begin{aligned} & \mathrm{proj}_{\mathbb{U}_d}\left(\mathrm{mat}\left(\left[\left(U_{\{t\}} \otimes I_n\right) + \left(I_n \otimes U_{\{t\}}\right)K\right]^T W \otimes W\left(U_{\{t\}} \otimes I_n\right) \quad \vec{U}_{\{t\}}\right)\right) \\ = \; & \mathrm{proj}_{\mathbb{U}_d}\left(\mathrm{mat}\left(\left(U_{\{t\}}^T \otimes I_n\right)W \otimes W\left(U_{\{t\}} \otimes I_n\right)\vec{U}_{\{t\}} \quad + \quad K^T\left(I_n \otimes U_{\{t\}}^T\right)W \otimes W\left(U_{\{t\}} \otimes I_n\right)\vec{U}_{\{t\}}\right)\right) \\ = \; & \mathrm{proj}_{\mathbb{U}_d}\left(\mathrm{mat}\left(\left(U_{\{t\}}^T W U_{\{t\}} \otimes W\right)\vec{U}_{\{t\}} \quad + \quad K^T\left(W U_{\{t\}} \otimes U_{\{t\}}^T W\right)\vec{U}_{\{t\}}\right)\right) \\ = \; & \mathrm{proj}_{\mathbb{U}_d}\left(\mathrm{mat}\left(\overrightarrow{W U_{\{t\}} U_{\{t\}}^T W U_{\{t\}}} \quad + \quad K^T \overrightarrow{U_{\{t\}}^T W U_{\{t\}} U_{\{t\}}^T W}\right)\right) \\ \overset{K^T\vec{A^T}=\vec{A}}{=} \; & \mathrm{proj}_{\mathbb{U}_d}\left(\mathrm{mat}\left(\overrightarrow{W U_{\{t\}} U_{\{t\}}^T W U_{\{t\}}} \quad + \quad \overrightarrow{W U_{\{t\}} U_{\{t\}}^T W U_{\{t\}}}\right)\right) \\ \overset{\text{Def. } W_{\{t\}}}{=} \; & \mathrm{proj}_{\mathbb{U}_d}\left(2\,\mathrm{mat}\left(\overrightarrow{W_{\{t\}} U_{\{t\}}}\right)\right) \\ \overset{\text{Lemma 4}}{=} \; & \mathrm{proj}_{\mathbb{U}_d}\left(\mathrm{mat}\left(\overrightarrow{W_{\{t\}} U_{\{t\}}}\right)\right) \\ = \; & \mathrm{proj}_{\mathbb{U}_d}\left(W_{\{t\}} U_{\{t\}}\right) \end{aligned}$$