MINIMUM-CORRECTION SECOND-MOMENT MATCHING:
THEORY, ALGORITHMS AND APPLICATIONS∗

JING LIN† AND PIERRE F. J. LERMUSIAUX †

Abstract. We address the problem of finding a matrix \( \tilde{U} \) as close to a given \( U \) as possible under the constraint that a prescribed second-moment matrix \( \tilde{P} \) must be matched, i.e. \( U^T \tilde{U} = \tilde{P} \). We develop the theory for this optimization problem by obtaining a closed-form formula for the unique global optimizer for the full-rank case, in both finite and infinite dimensions, and with an extension for rank deficient cases. We highlight the geometric intuition behind the theory and study the problem’s rich connections to minimum congruence transform, polar decomposition, optimal transport, and rank-deficient data assimilation. In the special case of \( \tilde{P} = I \), minimum-correction second-moment matching reduces to the well-studied optimal orthonormalization problem. We investigate the general strategies for numerically computing the optimizer and analyze existing polar decomposition and matrix square root algorithms. We then verify the higher performance of the various new algorithms using benchmark cases with randomly generated matrices. Lastly, we complete two applications for the stochastic Lorenz-96 differential equations in a chaotic regime. In reduced subspace tracking using dynamically orthogonal equations, we maintain the numerical orthonormality and continuity of time-varying base vectors. In ensemble square root filtering for data assimilation, the prior samples are transformed into unique posterior ones by matching the covariance given by the Kalman update while also minimizing the corrections to the prior samples.

Key words. Moment matching, minimum correction, optimal transport, polar decomposition, matrix square root, orthogonality constraint, re-orthonormalization, unitary integration, data assimilation, ensemble Kalman filter, ensemble square root filter.

AMS subject classifications. 65F25, 15A23, 60H15

1. Introduction. The second-moment matrix of \( U \in \mathcal{M}_{m \times n}(\mathbb{R}) \) defined as \( P = U^T \Gamma_m U \in \mathcal{M}_{n \times n} \) contains all the pairwise inner products of \( U \)'s columns weighted by \( \Gamma_m \). When we view \( U \)'s columns as vectors in an inner product space, \( P \) is called the Gram matrix. When the \( U \)'s rows are samples of a zero-mean random vector, \( P \) becomes the sample covariance matrix with \( \Gamma_m = (1/m)I \). Here we unify them as second-moment matrices. In many computational problems, we encounter the task of correcting a given matrix \( U \) to some \( \tilde{U} \) that matches a prescribed second-moment matrix \( \tilde{P} \), i.e. \( \tilde{P} = U^T \Gamma_m U \). However, such a \( \tilde{U} \) is not unique. When there is no physical information that favors one choice over another, a natural approach is to aim for a correction \( \tilde{U} - U \) that is minimal to avoid introducing numerical artifacts.

Precisely, given a target symmetric positive definite (SPD) second-moment matrix \( \tilde{P} \), as well as an inner product on \( \mathbb{R}^m \) and on \( \mathbb{R}^n \) defined by the SPD weight matrix \( \Gamma_m \in \mathcal{M}_{m \times m} \) and \( \Gamma_n \in \mathcal{M}_{n \times n} \), respectively, we want to solve the optimization

\[
\arg \min_{U \in \mathcal{M}_{m \times n} : U^T \Gamma_m U = \tilde{P}} \| U - \tilde{U} \|_{F, \Gamma_m, \Gamma_n}^2 , \tag{1}
\]

where \( \| \cdot \|_{F, \Gamma_m, \Gamma_n} \) is the Frobenius norm weighted by \( \Gamma_m \) and \( \Gamma_n \):

\[
\| V \|_{F, \Gamma_m, \Gamma_n} \overset{\Delta}{=} \text{tr}(\sqrt{\Gamma_m} V^T \Gamma_m V \sqrt{\Gamma_n}^T) = \text{tr}(\sqrt{\Gamma_m} V \sqrt{\Gamma_n}^T V^T) = \| V^T \|_{F, \Gamma_n, \Gamma_m} , \tag{2}
\]

∗Submitted to the editors XX/XX/2019.
†Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139 (linjing@mit.edu, pierrel@mit.edu).
‡Since we will exclusively focus on real matrices in this paper, the explicit specification “(\( \mathbb{R} \))” of the field of matrix elements will be dropped hereafter. However, everything discussed here readily generalizes to the complex case.