Polynomial perturbations of normal distributions

A. M. Davie

September 25, 2017

1 Introduction

In this paper we consider families of probability distributions on $\mathbb{R}^q$ of form $(\mathbb{P}_\epsilon)$ for ϵ small, where $\mathbb{P}_\epsilon$ has density f_ϵ with asymptotic expansion (in a sense that will be made precise)

$$f_\epsilon(x) \sim \phi_\Sigma(x)\left(1+\sum_{j=1}^{\infty}\epsilon^j S_j(x)\right) \tag{1}$$

where ϕ_Σ is the density of $N(0,\Sigma)$ and the S_j are polynomials on $\mathbb{R}^q$. Random variables with distributions of this type occur in various contexts, and the question which motivated this paper is the following: given a random variable X_ϵ with distribution of the form (1), which we wish to simulate but is hard to simulate directly, can we find find another random variable Y_ϵ which we can simulate more easily, and which has a distribution which is 'close' to that of X_ϵ, so that it act as an adequate substitute? The Vaserstein distances $\mathbb{W}_p$ seem to be suitable measures of 'closeness' of distributions, as they measure the L^p distance between X_ϵ and Y_ϵ when we use an optimal coupling between the two distributions. With this as motivation, we give estimates for Vaserstein distances between families of this type and methods for approximate simulation of random variables having distributions $\mathbb{P}_\epsilon$.

Our motivating example of such a family is given by stochastic Taylor expansions that arise in the numerical solution of SDEs driven by Brownian motion, as described for example in Chapter 5 of [5]; here we take ϵ to be $h^{1/2}$ where h is the stepsize of the numerical scheme. We have explored the use of coupling for the simulation of such expansions in [2], and the results of the present paper lead to an alternative treatment of some of the results of [2].

Another example is where $\mathbb{P}_\epsilon$, for $\epsilon = m^{-1/2}$ is the distribution of $m^{-1/2}(X_1 + \cdots X_m)$ where the X_i are i.i.d. $\mathbb{R}^q$-valued random variables with covariance Σ, and in this case we are able to obtain some fairly precise estimates for the multivariate CLT and approximation by Edgeworth expansions in Vaserstein metrics. We mention that this theory could potentially be extended to other situations where Edgeworth-type expansions of the form (1) are available, such as described for example in Chapter 2 of [3].

To explain the main idea of the method, fix $q \in \mathbb{N}$ and let P denote the space of all real-valued polynomials on $\mathbb{R}^q$, and P^q the space of $\mathbb{R}^q$-valued polynomial functions on $\mathbb{R}^q$. We also fix a positive-definite $q \times q$ matrix Σ. Let $p_1, \cdots, p_k \in P^q$. For $\epsilon \in \mathbb{R}$ we define $\rho_\epsilon : \mathbb{R}^q \to \mathbb{R}^q$ by $\rho_\epsilon(x) = x + \sum_{j=1}^k \epsilon^j p_j(x)$. We are interested in the distribution of $\rho_\epsilon(X)$ where X is an $\mathbb{R}^q$-valued random variable with $N(0,\Sigma)$ distribution and ϵ is close to 0. If we assume that ρ_ϵ is bijective, then this distribution has a density given by

$$f_\epsilon(y) = \det(D\rho_\epsilon^{-1}(y))\phi_\Sigma(\rho_\epsilon^{-1}(y)) \tag{2}$$

Bijectivity will generally only hold on some bounded region of $\mathbb{R}^q$, which will be large if ϵ is small; we will see that this is sufficient to use (2) to get an asymptotic expansion of the form(1). This then extends to the case where the polynomials p_j have random coefficients.

The idea then is to use such asymptotic expansions for densities to estimate Vaserstein distances between probability distributions of the described type. An important step in the argument is to show that the algebraic procedure for obtaining the polynomials (S_j) which appear in (1) from given (possibly random) polynomials (p_j) can be reversed - given an expansion (1) we can construct corresponding (deterministic) polynomials (p_j), which can then be used to give appropriate couplings.

In section 2 we describe the formal expansion of the density using (2) and some algebraic results that we shall need. The main results are given in section 3, and the applications to stochastic Taylor expansions and to Vaserstein versions of the CLT in section 4. Then in section 5 we describe how to obtain an asymptotic expansion in powers of ϵ for the $\mathbb{W}_2$ distance between two families of the form (1). In section 6 this is applied to the CLT and used to obtain a monotonicity result for $\mathbb{W}_2$ distances, related to a question of Villani [11].

We mention some notation that we shall use. If $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are probability measures on $\mathbb{R}^q$, then for $p \geq 1$ the Vaserstein distance $\mathbb{W}_p(\mathbb{P}, \tilde{\mathbb{P}})$ is defined as $\inf(\mathbb{E}|X - \tilde{X}|^p)^{1/p}$ where the inf is over all joint distributions on $\mathbb{R}^q \times \mathbb{R}^q$ for $(X, \tilde{X})$ which have marginals $\mathbb{P}$ and $\tilde{\mathbb{P}}$. We sometimes abuse notation and write this as $\mathbb{W}_p(X, \tilde{X})$ or $\mathbb{W}_p(X, \tilde{\mathbb{P}})$.

We sometimes use matrix notation, regarding elements of $\mathbb{R}^q$ as column vectors. For a scalar-valued function on $\mathbb{R}^q$ we use ∇f for its gradient vector; for $g : \mathbb{R}^q \to \mathbb{R}^q$ we use Dg for its derivative matrix and $\nabla . g$ for its divergence (which equals $\mathrm{tr}(Dg)$).

2 Formal expansion of density

This section is (almost) purely algebraic. We introduce the expansion of (2) as $\phi_\Sigma(y)$ times a formal power series in ϵ whose coefficients are polynomials in y. The sense in which this series represents the density is considered in the next section.

We start with a formal expansion of ρ_ϵ^{-1}. We define a sequence of polynomial functions $r_n : \mathbb{R} \times \mathbb{R}^q \to \mathbb{R}^q$ recursively as follows: first set $r_0(\epsilon, x) = \sum_{j=1}^k \epsilon^{j-1} p_j(x)$ and then, assuming r_n defined, define $r_{n+1}(\epsilon, x) = \epsilon^{-1}\{r_n(\epsilon, x) - r_n(0, \rho_\epsilon(x))\}$. Then define $u_n(x) = -r_{n-1}(0, x)$ for $n \in \mathbb{N}$, and also $\psi_n(\epsilon, x) = x + \sum_{j=1}^n \epsilon^j u_j(x)$. We see by induction on n that $\psi_n(\epsilon, \rho_\epsilon(x)) = x + \epsilon^{n+1} r_n(\epsilon, x)$. Then we can regard ψ_n as an approximation to ρ_ϵ^{-1} and $y + \sum_{j=1}^\infty \epsilon^j u_j(y)$ as a formal expansion of $\rho_\epsilon^{-1}(y)$.

The next step is to use ψ_n in place of ρ_ϵ^{-1} in (2). First define $\tau_n(\epsilon, y) = \sum_{j=1}^n \epsilon^j u_j(y)$, so that $\psi_n(y) = y + \tau_n(\epsilon, y)$, and then we see that $\phi_\Sigma(\psi_n(\epsilon, y)) = \phi_\Sigma(y) e^{\chi_n(\epsilon, y)}$ where

$$\chi_n(\epsilon, y) = -y^t \Sigma^{-1} \tau_n(\epsilon, y) - \frac{1}{2} \tau_n(\epsilon, y)^t \Sigma^{-1} \tau_n(\epsilon, y)$$

Then we truncate the exponential series for e^{χ_n} to get a polynomial $H_n(\epsilon, y) = \sum_{j=0}^n \frac{1}{j!} \chi_n(\epsilon, y)^j$, and use $H_n(\epsilon, y)\phi_\Sigma(y)$ as an approximation to $\phi_\Sigma(\rho^{-1}(y))$. We also use the polynomial $G_n(\epsilon, y) = \det(I + \sum_{j=1}^n \epsilon^j D u_j(y))$ as an approximation to $\det D\rho^{-1}(y)$. Finally we can write

$$H_n(\epsilon, y) G_n(\epsilon, y) = L_n(\epsilon, y) + \epsilon^{n+1} Q_n(\epsilon, y)$$

where Q_n is a polynomial on $\mathbb{R} \times \mathbb{R}^q$ and $L_n(\epsilon, y) = 1 + \sum_{j=1}^n \epsilon^j S_j(y)$ where $S_j \in P$ for each j. We will prove below that $L_n(\epsilon, y)\phi_\Sigma(y)$ approximates the density of $\rho_\epsilon(X)$ to order $O(\epsilon^{n+1})$ in a suitable sense.

We first make an observation about this construction. If we repeat the construction with m in place of n, where $m < n$, then we see that the polynomials ψ_m, H_m, G_m, L_m we obtain will be the same as ψ_n, H_n, G_n, L_n as far as terms in ϵ^m, since we only omit terms with higher powers of ϵ. It follows then that, for a given j, the polynomial S_j obtained above will be independent of n, as long as $n \geq j$. So we get a well-defined sequence $S_1, S_2, \cdots$ determined by $p_1, \cdots, p_k$.

Let P_Σ denote the subspace of $S \in P$ such that $\int_{\mathbb{R}^q} S(y)\phi_\Sigma(y)dy = 0$. This definition is motivated by the following lemma:

Lemma 1. *Let $S_1, S_2, \cdots$ be constructed from $p_1, \cdots, p_k$ as above. Then $S_j \in P_\Sigma$ for each $j \in \mathbb{N}$.*

The proof of this lemma is postponed to the next section. It follows from the characterisation of P_Σ below that the assertion of the lemma is purely algebraic, and it would not be hard to construct a direct algebraic proof. But it will be somewhat shorter to deduce it is a corollary of proposition 1 below.

We can characterise P_Σ as follows. Let $\mathcal{L}_\Sigma : P^q \to P$ be the linear mapping defined by $\mathcal{L}_\Sigma p(x) = x^t\Sigma^{-1}p(x) - \nabla.p(x)$. Then $\nabla.(\phi p)(x) = -\mathcal{L}_\Sigma p(x)\phi(x)$ and it follows from the divergence theorem that $\mathcal{L}_\Sigma p \in P_\Sigma$ for every $p \in P^q$. In the converse direction, a simple induction on the degree of u (see the proof of Lemma 1 in [2]) shows that if $u \in P_\Sigma$ then u is in the range of $\mathcal{L}_\Sigma$. So P_Σ is precisely the range of $\mathcal{L}_\Sigma$.

The above argument in fact shows that any element of P_Σ can be expressed as $\mathcal{L}_\Sigma\nabla u$ for some $u \in P$; moreover a similar induction on degree gives that if $r \in P$ and $\mathcal{L}_\Sigma\nabla r = 0$ then r is constant, so the u is unique up to an additive constant. Hence if we define P_G^q to be the the set of $p \in P^q$ of the form $p = \nabla u$ with $u \in P$, we have that $\mathcal{L}_\Sigma$ is bijective from $P_G^q \to P_\Sigma$, and we can define the inverse linear mapping $\mathcal{L}_\Sigma^{-1} : P_\Sigma \to P_G^q$.

Eigenfunctions of $\mathcal{L}_\Sigma\nabla$

We mention here that an explicit description of the action of $\mathcal{L}_\Sigma$ can be given in terms of Hermite polynomials, if we choose a coordinate system so that Σ is diagonal, with entries $\sigma_1^2, \cdots, \sigma_q^2$ where each $\sigma_j > 0$. Then if $u \in P$ is defined by $u(x) = \prod_{j=1}^q H_{m_j}(x_j/\sigma_j)$ where $m_1, \cdots, m_q$ are nonnegative integers, we have $\mathcal{L}_\Sigma\nabla u = \lambda u$ where $\lambda = \sum_{j=1}^q m_j\sigma_j^{-2}$. Now the set of such u, where the m_j are not all 0, spans P_Σ, and it follows that $\mathcal{L}_\Sigma\nabla$ maps P_Σ bijectively onto itself, from which one can again deduce that $\mathcal{L}_\Sigma$ maps P_G^q bijectively into P_Σ.

Returning to the construction of the S_j from the p_j, we observe that for $j < n$, S_j depends only on $\{p_m : m \le j\}$. It follows that, given a sequence $p_1, p_2, \cdots$ with $p_j \in P^q$ we obtain a well-defined sequence $S_1, S_2, \cdots$. Now we introduce the notation $\mathcal{P}$ for the set of all sequences $(u_1, u_2, \cdots)$ with $u_j \in P$, and similarly $\mathcal{P}^q$, $\mathcal{P}_\Sigma$ and $\mathcal{P}_G$. Then we define $\mathcal{S}_\Sigma : \mathcal{P}^q \to \mathcal{P}_\Sigma$ by $\mathcal{S}_\Sigma(p_1, p_2, \cdots) = (S_1, S_2, \cdots)$ as just described. We also write $\mathcal{S}_\Sigma^{(n)}$ for the truncated mapping: $\mathcal{S}_\Sigma^{(n)}(p_1, p_2, \cdots) = (S_1, S_2, \cdots, S_n)$.

Then we have the following:

Lemma 2. *The mapping $\mathcal{S}_\Sigma : \mathcal{P}_G^q \to \mathcal{P}_\Sigma$ is a bijection.*

Proof. Suppose $(S_1, S_2, \cdots) \in \mathcal{P}_\Sigma$. We have to show that there is a unique sequence $(p_1, p_2, \cdots)$ with $p_j \in P_G^q$ such that $\mathcal{S}_\Sigma(p_1, p_2, \cdots) = (S_1, S_2, \cdots)$. We do this by showing by induction on n that for each n there is a unique choice of $p_1, \cdots, p_n \in P_G^q$ such that $\mathcal{S}_\Sigma^{(n)}(p_1, \cdots, p_n) = (S_1, \cdots, S_n)$.

So we suppose we have such $p_1, \cdots, p_n$, and look for $p_{n+1} \in P_G^q$ such that $\mathcal{S}_\Sigma^{(n+1)}(p_1, \cdots, p_{n+1}) = (S_1, \cdots, S_{n+1})$. We have $\mathcal{S}_\Sigma^{(n+1)}(p_1, \cdots, p_n) = (S_1, \cdots, S_n, v)$ where $v \in P_\Sigma$. Then for any choice of $p_{n+1} \in P_G^q$, we have $\mathcal{S}_\Sigma^{(n+1)}(p_1, \cdots, p_{n+1}) = (S_1, \cdots, S_n, v + \mathcal{L}_\Sigma p_{n+1})$. So we require $\mathcal{L}_\Sigma p_{n+1} = S_{n+1} - v$, and by the bijectivity of $\mathcal{L}_\Sigma$ there is a unique such $p_{n+1} \in P_G^q$. This completes the inductive step. The initial step is proved in the same way. □

We remark that if $q > 1$ then any element of $\mathcal{P}_\Sigma$ will have many preimages under $\mathcal{S}_\Sigma$ in $\mathcal{P}^q$, and for the results in Section 3 any reasonable way of choosing a preimage would work. But for $\mathbb{W}_2$ bounds the preimage in P_G^q has an optimality property, summarised in Lemma

15, which is important in Sections 5 and 6.

Some special cases. To illustrate the action of $\mathcal{S}$, we construct the first 3 terms in the case $q=1$, assuming $\Sigma = 1$ for simplicity. We first calculate the polynomials u_1, u_2, u_3 that appear in the formal expansion of ρ_ϵ, obtaining $u_1(y) = -p_1(y)$, $u_2(y) = p_1'(y)p_1(y) - p_2(y)$, and

$$u_3(y) = p_1'(y)p_2(y) + p_1(y)p_2'(y) - p_1'(y)^2 p_1(y) - \frac{1}{2}p_1''(y)p_1(y)^2 - p_3(y)$$

Using this, we find after some calculation that $S_1(y) = yp_1(y) - p_1'(y)$,
$S_2(y) = \frac{1}{2}(y^2-1)p_1(y)^2 + p_1''(y)p_1(y) + p_1'(y)^2 - p_2'(y) + y\{p_2(y) - 2p_1'(y)p_1(y)\}$ and

$$\begin{aligned} S_3(y) =& y\left(\frac{3}{2}p_1''(y)p_1(y)^2 + 3p_1'(y)^2 - 2p_1'(y)p_2(y) - 2p_1(y)p_2'(y) + p_3(y)\right) \\ &+ (1-y^2)\left(\frac{3}{2}p_1'(y)p_1(y)^2 - p_1(y)p_2(y)\right) + \frac{1}{6}(y^3-3y)p_1(y)^3 + p_1''(y)p_2(y) + 2p_1'(y)p_2'(y) \\ &+ p_1(y)p_2''(y) - 3p_1''(y)p_1'(y)p_1(y) - \frac{1}{2}p_1'''(y)p_1(y)^2 - p_1'(y)^3 - p_3'(y) \end{aligned}$$

We next give the expressions for S_1 and S_2 for general q and Σ; these are similar to those above but involve partial derivatives. We find $S_1(y) = y^t\Sigma^{-1}p_1(y) - \nabla.p_1(y)$ and

$$\begin{aligned} S_2(y) =& \frac{1}{2}(y^t\Sigma^{-1}p_1(y))^2 - y^t\Sigma^{-1}p_1(y)\nabla.p_1(y) - \frac{1}{2}p_1(y)\Sigma^{-1}p_1(y) - y^t\Sigma^{-1}Dp_1(y)p_1(y) \\ &+ \frac{1}{2}(\nabla.p_1(y))^2 + \frac{1}{2}\text{tr}(Dp_1(y)^2) + p_1(y)^t\nabla(\nabla.p_1)(y) + y^t\Sigma^{-1}p_2(y) - \nabla.p_2(y) \end{aligned}$$

We remark that in the case $\rho_\epsilon(x) = x + \epsilon Ax$, where A is a fixed $q \times q$ matrix, we have $p_1(x) = Ax$ and $p_2 = 0$ and we find that $S_1(y) = y^t\Sigma^{-1}Ay - \text{tr}A$ and

$$S_2(y) = \frac{1}{2}(y^t\Sigma^{-1}Ay)^2 - (\text{tr}A)y^t\Sigma^{-1}Ay - \frac{1}{2}y^tA^t\Sigma^{-1}Ay - y^t\Sigma^{-1}A^2y + \frac{1}{2}(\text{tr}A)^2 + \frac{1}{2}\text{tr}(A^2)$$

which in the case $\Sigma = I$ gives the conclusion of Lemma 8 of [2]. Note that in this case $\rho_\epsilon(X)$ has $N(0,\Sigma_\epsilon)$ distribution, where $\Sigma_\epsilon = (I+\epsilon A)\Sigma(I+\epsilon A^t)$, and $1 + \epsilon S_1(y) + \epsilon^2 S_2(y)$ is simply the expansion to order ϵ^2 of $f_\epsilon(y)\phi_\Sigma(y)$ where f_ϵ is the $N(0,\Sigma_\epsilon)$ density function.

A similar special case is $\rho_\epsilon(x) = x + \epsilon v$, where $v \in \mathbb{R}^q$ is fixed. Then we have $p_1(x) = v$, $p_2 = 0$ and all derivatives of p_1 are zero. We find $S_1(y) = v^t\Sigma^{-1}y$ and $S_2(y) = \frac{1}{2}\{(v^t\Sigma^{-1}y)^2 - vt\Sigma^{-1}v\}$. In this case ρ_ϵ has $N(\epsilon v, \Sigma)$ distribution and the series $1 + \epsilon S_1(y) + \epsilon^2 S_2(y) + \cdots$ is just the expansion in powers of ϵ of $\exp(\epsilon v^t\Sigma^{-1}y - \frac{1}{2}\epsilon^2 v^t\Sigma^{-1}v)$, which (exceptionally) converges for all ϵ and y.

We remark that, in the case $q = 1$, taking $\Sigma = 1$ again, the correspondence between $\mathcal{P}$ (which is the same as $\mathcal{P}_G^q$ in this case) and $\mathcal{P}_1$ given by $\mathcal{S}_1$ is equivalent to the correspondence between a Cornish-Fisher expansion and an Edgeworth expansion for the corresponding density. For if $\mathbb{P}_\epsilon$ has distribution function F_ϵ, and density with asymptotic expansion (1), and $X \sim N(0,1)$, then the random variable $F_\epsilon^{-1} \circ \Phi(X)$ will have distribution $\mathbb{P}_\epsilon$, and the series $x + \sum \epsilon^j p_j(x)$, which is the formal expansion of $F_\epsilon^{-1} \circ \Phi$, is the Cornish-Fisher expansion corresponding to the Edgeworth expansion (1). There is an extensive literature on these expansions, which are mainly used in Statistics - see for example Chapter 2 of [3] or Chapter 6 of [4]. (In statistical applications, usually $\epsilon = m^{-1/2}$ where m is a sample size).

3 Main results

The following proposition shows that, in a suitable sense, the series $(1 + \sum_{j=1}^\infty \epsilon^j S_j)\phi_\Sigma$ can be regarded as an asymptotic expansion for the density of $\rho_\epsilon(X)$.

Proposition 1. *With notation as above, let $\mathbb{P}_\epsilon$ be the probability distribution of $\rho_\epsilon(X)$ and let $\nu_{\epsilon,n}$ be the signed measure on $\mathbb{R}^q$ with density $\phi_\Sigma(y)L_n(\epsilon,y)$. Then for any $M \geq 1$ we have a bound*

$$\int_{\mathbb{R}^q}(1+|y|)^M d|\mathbb{P}_\epsilon - \nu_{\epsilon,n}|(y) \leq CK^N|\epsilon|^{n+1} \tag{3}$$

for $\epsilon \in [-1,1]$, where C and N are positive constants depending only on k,n,q,M and the maximum degree d of $p_1,\cdots,p_k$, and $K \geq 1$ is an upper bound for the absolute values of the coefficients of $p_1,\cdots,p_k$ and for $\|\Sigma\|$ and $\|\Sigma^{-1}\|$.

Proof. We first prove the proposition in the case where $\Sigma = I$ so that $\phi_\Sigma = \phi$, the $N(0,I)$ density function.

We use C_1, C_2 etc to denote positive constants which depend only on k,n,q,M,d. First we can find $C_1 \geq 1$ such that

$$\max(|\chi_n(\epsilon,\rho_\epsilon(x))|,\ |\tau_n(\epsilon,\rho_\epsilon(x))|,\ |\rho_\epsilon(x)-x|,\ \|D\rho_\epsilon(x)-I\|) \leq C_1|\epsilon|K^{C_1}(1+|x|)^{C_1} \tag{4}$$

and

$$\max(|r_n(\epsilon,x)|,\ \|Dr_n(\epsilon,x)\|,\ |Q_n(\epsilon,x)|) \leq C_1K^{C_1}(1+|x|)^{C_1} \tag{5}$$

for all $x \in \mathbb{R}^q$. Then let $R = (2C_1K)^{-1}|\epsilon|^{-1/(2C_1)} - 1$ and let $B_R = \{x \in \mathbb{R}^q : |x| < B_R\}$ (which will of course be empty if $R \leq 0$, which can happen if ϵ is not very small).

Now define a measure μ_ϵ as the image under ρ_ϵ of the restriction to B_R of the $N(0,I)$ distribution on $\mathbb{R}^q$. We also define $\tilde{\nu}_\epsilon = \nu|\rho_\epsilon(B_R)$. Then we have

$$\int_{\mathbb{R}^q}(1+|y|)^M d|\mathbb{P}_\epsilon - \nu_{\epsilon,n}|(y) = \Omega_1+\Omega_2+\Omega_3$$

where $\Omega_1 = \int_{\mathbb{R}^q}(1+|y|)^M d|\mu_\epsilon - \tilde{\nu}_\epsilon|(y)$, $\Omega_2 = \int_{\mathbb{R}^q}(1+|y|)^M d(\mathbb{P}_\epsilon - \mu_\epsilon)(y)$ and $\Omega_3 = \int_{\mathbb{R}^q}(1+|y|)^M d|\nu_\epsilon - \tilde{\nu}_\epsilon|(y)$.

We first bound Ω_1. To this end we note that, by the definition of R, for $x \in B_R$ the RHS of (4) is bounded by $\frac{1}{2}|\epsilon|^{1/2}$. It then follows from (4) that for $x \in B_R$ we have $\|D\rho_\epsilon(x) - I\| \leq \frac{1}{2}$ and so ρ_ϵ is bijective on B_R. Then we have

$$\Omega_1 = \int_{\rho_\epsilon(B_R)}(1+|y|)^M|\det D\rho_\epsilon^{-1}(y)\phi(\rho_\epsilon^{-1}(y)) - L_n(y)\phi)|dy$$

To bound the RHS, we fix $x \in B_R$ and set $y = \rho_\epsilon(x)$, noting that $|x-y| \leq 1$ by (4). Then, noting that $G_n(\epsilon,y)\det D\rho_\epsilon(x) = \det(I+\epsilon^{n+1}Dr_n(\epsilon,y)) = 1+\epsilon^{n+1}s_n(\epsilon,y)$ where s_n is a polynomial, we see that

$$\Omega_1 = \int_{\rho_\epsilon(B_R)}(1+|y|)^M\left|\det D\rho_\epsilon^{-1}(y)\{\phi(x) - H_n(\epsilon,y)\phi(y)(1+\epsilon^{n+1}s_n(\epsilon,y))\} + \epsilon^{n+1}Q_n(\epsilon,y)\phi(y)\right|dy \tag{6}$$

To bound $\phi(x) - H_n(\epsilon,y)\phi(y)$, note first that by (4), $|x-y|$ and $|\psi_n(\epsilon,y)-y| = |\tau_n(\epsilon,y)|$ are both $\leq |\epsilon|1/2 \leq \min(1,|y|^{-1})$, the last inequality following from the fact that $|y| \leq 1+R \leq |\epsilon|^{-1/2}$, and so, for any z on the straight line segment joining x to $\psi(y)$, we have $\phi(z) \leq e\phi(y)$ and so $|D\phi(z)| = |z|\phi(z) \leq e(1+|y|)\phi(y)$. We then deduce that

$$|\phi(x) - \phi(\psi_n(\epsilon,y)| \leq |\epsilon|^{n+1}e(1+|y|)|r_n(\epsilon,x)|\phi(y)$$

Also we have $\phi(\psi_n(\epsilon,y)) = \phi(y)e^{\chi_n(\epsilon,y)}$ and, since $|\chi_n(\epsilon,y)| \leq 1$, we have $|e^{\chi_n(\epsilon,y)} - H_n(y)| \leq |\chi_n(\epsilon,y)|^{n+1}$. Putting these bounds together we have

$$|\phi(x) - H_n(\epsilon,y)\phi(y)| \leq C_2K^{C_2}(1+|y|)^{C_2}|\epsilon|^{n+1}\phi(y)$$

Now since $\|D\rho_\epsilon(x)-I\| \le 1/2$ we have $|\det D\rho^{-1}(y)| = |\det D\rho(x)|^{-1} \le 2^q$, and then, using (5) to bound the ϵ^{n+1} terms in (6), we obtain

$$\Omega_1 \le C_3 K^{C_3}|\epsilon|^{n+1}\int_{\rho_\epsilon(B_R)}(1+|y|)^{C_3}\phi(y)dy$$

from which we get the required bound for Ω_1.

It remains to bound Ω_2 and Ω_3. We have $\Omega_2 = \int_{B_R^c}(1+|\rho_\epsilon(x)|)^M\phi(x)dx$. We can find a constant C_4 such that $\phi(x) \le C_4(1+|x|)^{-C(n+1)-M-d-q-1}$ for $x \in \mathbb{R}^q$. Then for $x \in B_R^c$ we have $1+|x| > R$ so $\phi(x) \le C_4 R^{-C(n+1)}(1+|x|)^{-M-d-q-1} \le C_5\epsilon^{n+1}(1+|x|)^{-M-d-q-1}$, using the definition of R for the last inequality. It follows that Ω_2 satisfies an inequality of the required form. A similar argument applies to $\Omega_3 = \int_{\rho(B_R)^c}(1+|y|)^M|L_n(y)|\phi(y)dy$, completing the proof for the case $\Sigma = I$.

In the case of general positive definite Σ we can write $X = AX^*$ where $A = \Sigma^{1/2}$ and X^* is $N(0,I)$, and then we can write $p_j(X) = Ap_j^*(X^*)$ where the p_j^* are again polynomials. Then $\rho_\epsilon(X) = A\rho_\epsilon^*(X^*)$ where $\rho_\epsilon^*(x) = x + \sum_{j=1}^k \epsilon^j p_j^*(x)$, and the case just proved applies to $\rho_\epsilon^*(X^*)$. The required result follows, on noting the the coefficients of the p_j^* are bounded by a suitable power of K. □

An immediate consequence of Proposition 1 is that $\int_{\mathbb{R}^q} L_n(\epsilon,y)\phi_\Sigma(y)dy = 1 + O(|\epsilon|^{n+1})$ as $\epsilon \to 0$, which implies that $\int_{\mathbb{R}^q} S_j(y)\phi_\Sigma(y)dy = 0$, proving Lemma 1.

We now define a general class of families of distributions, of which Proposition 1 gives examples.

Definition 1. *Let $E \subseteq [-1,1]\backslash\{0\}$ with $0 \in \overline{E}$ and suppose $(\mathbb{P}_\epsilon : \epsilon \in E)$ is a family of probability measures on $\mathbb{R}^q$. We say $(S_1, S_2, \cdots) \in \mathcal{P}_\Sigma$ is an $\mathcal{A}_\Sigma$-sequence for the family $(\mathbb{P}_\epsilon)$ if, for every choice of $n \in \mathbb{N}$ and $M \ge 1$, we can find $C > 0$ such that for every $\epsilon \in E$ we can find a probability measure θ_ϵ supported on $\{x \in \mathbb{R}^q : |x| < |\epsilon|^{n+1}\}$ such that*

$$\int_{\mathbb{R}^q}(1+|y|)^M d|\mathbb{P}_\epsilon * \theta_\epsilon - \nu_{\epsilon,n}|(y) \le C|\epsilon|^{n+1} \tag{7}$$

where $*$ denotes convolution and $\nu_{\epsilon,n}$ is the measure on $\mathbb{R}^q$ with density $\phi_\Sigma(y)(1+\sum_{j=1}^n \epsilon^j S_j(y))$.

We remark that the convolution with θ_ϵ is included in order to allow some cases where $\mathbb{P}_\epsilon$ is singular. In many applications it can be omitted (which is equivalent to taking θ_ϵ to be the point mass at 0).

It follows from Proposition 1 that if $p_1, \cdots, p_k \in P^q$ then $\mathcal{S}(p_1, \cdots, p_n)$ is an $\mathcal{A}_\Sigma$-sequence for the family given by the distributions of $\rho_\epsilon(X)$ as described above. We now extend this to the case where $p_1, \cdots, p_k$ are random polynomials, independent of X. More precisely, we have the following:

Lemma 3. *Suppose that the $p_1, \cdots, p_k \in P^q$ are polynomials of fixed degrees whose coefficients are random variables having finite moments of all orders, and that $X \sim N(0,\Sigma)$ is independent of this collection of random variables. For $\epsilon \in [-1,1]\backslash\{0\}$ let $\mathbb{P}_\epsilon$ be the distribution of the random variable $\rho_\epsilon(X) = X + \sum_{j=1}^k p_j(X)$. Let $(S_1, S_2, \cdots) = \mathcal{S}_\Sigma(p_1, p_2, \cdots)$, which is a sequence of random polynomials in P_0.*

Then the family $(\mathbb{P}_\epsilon)$ has $(\overline{S}_1, \overline{S}_2, \cdots)$ as an $\mathcal{A}_\Sigma$-sequence, where $\overline{S}_j(y) = \mathbb{E}S_j(y)$.

Proof. We let $\mathbb{P}_\epsilon^c$ denote the density of $\rho_\epsilon(X)$, conditional on the coefficients; this is then a random measure. Also let $\nu_{\epsilon,n}^c$ be the (random) measure with density $\phi_\Sigma(y)(1+\sum_{j=1}^n \epsilon^j S_j(y))$. From (3) we have

$$\int_{\mathbb{R}^q}(1+|y|)^M d|\mathbb{P}_\epsilon^c - \nu_{\epsilon,n}^c|(y) \le CK^N|\epsilon|^{n+1} \tag{8}$$

where, for given n and M, C and N are fixed, and K is a random variable with all moments finite.

Now note that $\mathbb{P}_\epsilon = \mathbb{E}\mathbb{P}^c_\epsilon$, and let $\nu_{n,\epsilon} = \mathbb{E}\nu^c_{n,\epsilon}$, which can be expressed as the measure with density $\phi(y)(1+\sum_{j=1}^k \epsilon^j \overline{S}_j(y))$. Then taking expectation and using (8) gives, for given M and n,

$$\int_{\mathbb{R}^q} (1+|y|)^M d|\mathbb{P}_\epsilon - \nu_{\epsilon,n}|(y) \leq \mathbb{E}\int_{\mathbb{R}^q} (1+|y|)^M d|\mathbb{P}^c_\epsilon - \nu^c_{\epsilon,n}|(y) \leq C\mathbb{E}(K^N)|\epsilon|^{n+1} \tag{9}$$

from which the result follows (taking θ_ϵ to be the point mass at 0). □

Example. We suppose $q = 1$ and define $\rho_\epsilon(x) = x + \epsilon\alpha(x+z)$ where $z \in \mathbb{R}$ is fixed and α is random, taking values ± 1 with probability $\frac{1}{2}$ each. We apply the above to $\rho_\epsilon(X)$ where X is $N(0,1)$, independent of α. We obtain $S_1(y) = y\alpha(y+z) - \alpha$ and $S_2(y) = \frac{1}{2}(y^2-1)(y+z)^2 + 1 - 2y(y+z)$, using the fact that $\alpha^2 = 1$. Then $\overline{S}_1(y) = 0$ and $\overline{S}_2(y) = \frac{1}{2}(y^2-1)(y+z)^2 + 1 - 2y(y+z)$. This example occurs (with somewhat different notation) in the treatment of the 'exact two-dimensional coupling' in section 8 of [2].

We now state a theorem giving bounds for Vaserstein distances.

Theorem 4. *Suppose $(\mathbb{P}_\epsilon : \epsilon \in E)$ and $(\tilde{\mathbb{P}}_\epsilon : \epsilon \in E)$ are families of probability distributions on $\mathbb{R}^q$ having respectively an $\mathcal{A}_\Sigma$-sequence $(S_1, S_2, \cdots)$ and an $\mathcal{A}_\Sigma$-sequence $(\tilde{S}_1, \tilde{S}_2, \cdots)$. Suppose also that, for some $n \in \mathbb{N}$, we have $S_j = \tilde{S}_j$ for $1 \leq j \leq n$. Let $M \geq 1$ be given. Then we can find $C > 0$ such that $\mathbb{W}_M(\mathbb{P}_\epsilon, \tilde{\mathbb{P}}_\epsilon) \leq C|\epsilon|^{n+1}$ for all $\epsilon \in E$.*

Proof. Let $(p_1, p_2, \cdots) = \mathcal{S}_\Sigma^{-1}(S_1, S_2, \cdots)$ and $(\tilde{p}_1, \tilde{p}_2, \cdots) = \mathcal{S}_\Sigma^{-1}(\tilde{S}_1, \tilde{S}_2, \cdots)$, noting that then $p_j = \tilde{p}_j$ for $1 \leq j \leq n$. Choose $r \in \mathbb{N}$ such that $r + 1 \geq M(n+1)$. Let X be an $N(0, \Sigma)$ random vector and let $Y_\epsilon = X + \sum_{j=1}^r \epsilon^j p_j(X)$ and $\tilde{Y}_\epsilon = X + \sum_{j=1}^r \epsilon^j \tilde{p}_j(X)$. Then let $\mathbb{Q}_\epsilon$ and $\tilde{\mathbb{Q}}_\epsilon$ be the distribution measures of Y_ϵ and $\tilde{Y}_\epsilon$ respectively.

Let $\nu_{\epsilon,r}$ be the measure with density $\phi_\Sigma(y)(1+\sum_{j=1}^r \epsilon^j S_j)$. Then $\int_{\mathbb{R}^q}(1+|y|)^M d|\mathbb{P}_\epsilon * \theta_\epsilon - \nu_{\epsilon,r}|(y) \leq C_1|\epsilon|^{r+1}$ and $\int_{\mathbb{R}^q}(1+|y|)^M d|\mathbb{Q}_\epsilon - \nu_{\epsilon,r}|(y) \leq C_1|\epsilon|^{r+1}$ so

$$\int_{\mathbb{R}^q} (1+|y|)^M d|\mathbb{P}_\epsilon * \theta_\epsilon - \mathbb{Q}_\epsilon|(y) \leq 2C_1|\epsilon|^{r+1}$$

It then follows from equation (11) in [2] (or proposition 7.10 in [11]) that $\mathbb{W}_M(\mathbb{P}_\epsilon * \theta_\epsilon, \mathbb{Q}_\epsilon) \leq C_2\epsilon^{(r+1)/M} \leq C_2|\epsilon|^{n+1}$. And $\mathbb{W}_M(\mathbb{P}_\epsilon * \theta_\epsilon, \mathbb{P}_\epsilon) \leq |\epsilon|^{n+1}$, so $\mathbb{W}_M(\mathbb{P}_\epsilon, \mathbb{Q}_\epsilon) \leq (1+C_2)|\epsilon|^{n+1}$. Similarly we have $\mathbb{W}_M(\tilde{\mathbb{P}}_\epsilon, \tilde{\mathbb{Q}}_\epsilon) \leq C_3|\epsilon|^{n+1}$.

Also we have

$$\mathbb{W}_M(\mathbb{Q}_\epsilon, \tilde{\mathbb{Q}}_\epsilon) \leq (\mathbb{E}|Y_\epsilon - \tilde{Y}_\epsilon|^M)^{1/M} = \left(\mathbb{E}\left|\sum_{j=n+1}^r \epsilon^j (p_j(X) - \tilde{p}_j(X))\right|^M\right)^{1/M} \leq C_4|\epsilon|^{n+1}$$

The result then follows from the triangle inequality. □

4 Applications

Stochastic Taylor expansion.

Lemma 3 applies to the stochastic Taylor expansion, in powers of $\epsilon = h^{1/2}$, of the solution at time h of an SDE $dx(t) = a(t, x(t))dt + B(t, x(t))dW(t)$ with $x(0) = x^{(0)}$ where $x(t) \in \mathbb{R}^q$, $a(t, x) \in \mathbb{R}^2$, $B(t, x)$ is a $q \times d$ matrix and W is a d-dimensional standard Brownian motion. In Theorem 4 of [2] one considers, for a fixed $m \in \mathbb{N}$, a random vector Z defined by $Z_i =$

$X_i+\hat{Q}_i(\epsilon, X, (\epsilon^{l(\alpha)-1}K_\alpha)_{\alpha\in M_m})$ where X has $N(0,\Sigma)$ distribution, $\hat{Q}$ is a polynomial obtained from the stochastic Taylor expansion and the K_α are iterated integrals of a Brownian bridge process. Here M_m denotes the set of all multi-indices $\alpha = (j_1, \cdots, j_l)$ of length $l = l(\alpha) \in \{2, 3, \cdots, m\}$.

The assertion of Theorem 4 of [2] is that, if one defines a modified random vector $\tilde{Z}$ by replacing K_α by L_α in the definition of$Z$, where the random variables L_α have all moments finite and are such that $\mathbb{E}(K_{\alpha_1}\cdots K_{\alpha_r}) = \mathbb{E}(L_{\alpha_1}\cdots L_{\alpha_r})$ whenever $\alpha_1, \cdots, \alpha_r \in \mathcal{M}_m$ satisfy $\sum_{k=1}^r (l(\alpha_k)-1) \le m-1$, then we have $\mathbb{W}_2(Z, \tilde{Z}) = O(\epsilon^m)$.

We now show that this theorem is a corollary of Theorem 4 above. We apply Theorem 4 with $\mathbb{P}_\epsilon$ and $\tilde{\mathbb{P}}_\epsilon$ being the distributions of Z and $\tilde{Z}$ respectively, and $n = m-1$. Using Lemma 3 we see that $\mathbb{P}_\epsilon$ has a $\mathcal{A}_\Sigma$-sequence $(S_1, S_2, \cdots)$ and $\tilde{\mathbb{P}}_\epsilon$ has a $\mathcal{A}_\Sigma$-sequence $(\tilde{S}_1, \tilde{S}_2, \cdots)$. The above moment hypothesis on the L_α then ensures that $S_j = \tilde{S}_j$ for $1 \le j \le m-1$, and Theorem 4 shows that $\mathbb{W}_2(\mathbb{P}_\epsilon, \tilde{\mathbb{P}}_\epsilon) \le C\epsilon^m$, which implies Theorem 4 of [2].

As an alternative way of applying Theorem 4, suppose we are given a sequence $(p_1, p_1, \cdots)$ of random polynomials as in the theorem, with a resulting sequence $(\overline{S}_1, \overline{S}_2, \cdots)$, and the define the (deterministic) sequence $(\tilde{p}_1, \tilde{p}_2, \cdots) = \mathcal{S}^{-1}(\overline{S}_1, \overline{S}_2, \cdots)$. Then Theorem 4 implies that the $\mathbb{W}_M$ distance between $X + \sum_{j=1}^n \epsilon^j p_j(X)$ and $X + \sum_{j=1}^n \epsilon^j \tilde{p}_j(X)$ is $O(\epsilon^{n+1})$. We can then apply this to the situation of Theorem 4 of [2], and instead of the approximation $\tilde{Z}$ constructed using additional random variables L_α, we can use a polynomial in the normal random vector X and no additional random variable, and still obtain a scheme of order $\frac{m}{2}$. Though it requires fewer random variables, this scheme does involve some added algebraic calculation, and is not necessarily simpler overall than the method using the L_α.

We also remark that the 'approximate coupling' method described in section 8 of [2] can be treated as an application of Theorem 4 to $\rho_\epsilon(U)$ where U is $N(0, I)$ and the random vector polynomial ρ_ϵ is defined by $\rho_\epsilon(x)_i = x_i + \epsilon\sum_{k,l=1}^d \sigma_{ikl}x_k U_l^* + \epsilon^2 p_i(x)$ and the polynomial vector p is chosen so that the resulting $\overline{S}_2$ is zero. As the distribution of $\rho_\epsilon(U)$ is an even function of ϵ, $\overline{S}_j$ is zero for odd j, and then it follows from Theorem 4 that the $\mathbb{W}_2$ distance between the distribution of $\rho_\epsilon(U)$ and $N(0, I)$ is $O(\epsilon^4)$. This approach is somewhat more rigorous than the argument given in [2].

Central Limit bounds in Vaserstein metrics.

Let $X_1, X_2, \cdots$ be i.i.d. $\mathbb{R}^q$-valued random variables and let $Y_m = m^{-1/2}(X_1 + \cdots + X_m)$. As another application of Theorem 4, we show that, under suitable conditions, the Edgeworth expansion gives an asymptotic expansion of the distribution of Y_m in $\mathbb{W}_M$ metric for any M. In particular, we obtain bounds for the normal approximation to Y_m given by the CLT.

We start by reviewing the standard derivation of the Edgeworth expansion, as can be found for example in [1] or [6]. We assume that $\mathbb{E}X_1 = 0$, as can be ensured by adding a constant vector, and that X_1 has nonsingular covariance matrix Σ. We also suppose that all moments of X_1 are finite. Let χ denote the characteristic function of X_1. Then near 0 we have the Taylor expansion $\log\chi(s) \sim -\frac{1}{2}s^t\Sigma s + \sum_{|\alpha|\ge 3} c_\alpha s^\alpha$. This expansion may or may not converge, but is asymptotic in the sense that, for each integer $n \ge 3$ there is a constant C_n such that

$$\left|\log\chi(s) + \frac{1}{2}s^t\Sigma s - \sum_{3\le|\alpha|\le n} c_\alpha s^\alpha\right| \le C_n|s|^{n+1} \tag{10}$$

for s near 0.

Now let ψ_m be the characteristic function of Y_m. Then $\log\psi_m(z) = m\log\chi(m^{-1/2}z) \sim -\frac{1}{2}z^t\Sigma z + \sum_{\alpha\ge 3} m^{1-|\alpha|/2}c_\alpha z^\alpha$. We can then write formally $\psi_m(z) \sim e^{-\frac{1}{2}z^t\Sigma z}(1 + \sum_{k=1}^\infty m^{-k/2}P_k(z))$ where $1 + \sum_{k=1}^\infty m^{-k/2}P_k(z)$ is the formal expansion of $\exp(\sum_{|\alpha|\ge 3} m^{1-\frac{|\alpha|}{2}}c_\alpha z^\alpha)$ in powers of $m^{-1/2}$. Here P_k is a polynomial with degree $3k$.

Next we recall that ψ_m is the Fourier transform of the density f_m of Y_m, and taking inverse transforms we obtain the expansion $f_m(x) \sim \phi_\Sigma(x)(1+\sum_{k=1}^\infty m^{-k/2}Q_k(x))$ where Q_k is a polynomial of degree $3k$. This is the Edgeworth expansion for the density f_m.

So far this discussion has been purely formal; without further conditions we cannot even say that Y_n has a density. We impose the following condition, sometimes known as the Cramer condition, and which we denote by CC: we say that the distribution of X_1 satisfies CC if the characteristic function χ satisfies $\limsup_{|s|\to\infty}|\chi(s)| < 1$. We note that CC is satisfied if X_1 has a density, or indeed if its distribution is not singular w.r.t. Lebesgue measure. It is also satisfied for some singular measures, such as the canonical measure on the Cantor middle-third set (in one dimension). We note also that if $|\chi(s)| = 1$ for some $s \in \mathbb{R}^q$, then $|\chi(ns)| = 1$ for all $n \in \mathbb{N}$. From this it follows that CC implies $|\chi(s)| < 1$ for all non-zero s.

Then we have the following:

Proposition 2. *With the notations and assumptions as above, let $E = \{m^{-1/2} : m \in \mathbb{N}\}$, and for $\epsilon = m^{-1/2}$ let $\mathbb{P}_\epsilon$ be the distribution of Y_m. Then $(Q_1, Q_2, \cdots)$ is an $\mathcal{A}$-sequence for the family $(\mathbb{P}_\epsilon)$.*

Proof. We fix $n \in \mathbb{N}$, and use $C_1, C_2, \cdots$ to denote constants which may depend on n but not on $\epsilon \in E$. We fix a smooth non-negative function h on $\mathbb{R}^q$, vanishing outside $\{x : |x| < 1\}$, such that $\int h d\lambda = 1$ where λ is Lebesgue measure on $\mathbb{R}^q$. Then for $\epsilon \in E$ we define $\theta_\epsilon(x) = \epsilon^{-q(n+1)}h(\epsilon^{-n-1}x)$, noting that θ_ϵ is a probability density supported on $\{x : |x| < \epsilon^{n+1}\}$.

Now we can find $\delta > 0$ so that (10) holds and $\log|\chi(s)| \le -\frac{1}{4}s^t\Sigma s$ whenever $|s| < \delta$. Then from the CC condition, together with the continuity of χ and the fact that $|\chi(s)| < 1$ for nonzero s, we can find $\gamma \in (0, 1)$ such that $|\chi(s)| < \gamma$ whenever $|s| \ge \delta$. Then $|\psi_m(z)| \le \exp(-\frac{1}{4}z^t\Sigma z)$ whenever $|z| \le m^{1/2}\delta$, and $|\psi_m(z)| \le \gamma^m$ whenever $|z| \ge m^{1/2}\delta$.

We also have

$$\left|\log\psi_m(z) + \frac{1}{2}z^t\Sigma z - \sum_{3\le|\alpha|\le n} m^{1-|\alpha|/2}c_\alpha z^\alpha\right| \le C_1 m^{(1-n)/2}|z|^{n+1}$$

for $|z| \le m^{1/2}\delta$, and for $|z| \le m^{1/6}$ we have

$$\left|\exp\left(\sum_{3\le|\alpha|\le n} m^{1-|\alpha|/2}c_\alpha z^\alpha\right) - 1 - \sum_{k=1}^{n-2} m^{-k/2}P_k(z)\right| \le C_2 m^{(1-n)/2}(1+|z|^{3n})$$

and combining these inequalities gives

$$\left|\psi_m(z) - e^{-\frac{1}{2}z^t\Sigma z}\left(1 + \sum_{k=1}^{n-2} m^{-k/2}P_k(z)\right)\right| \le C_3 m^{(1-n)/2}(1+|z|^{3n})e^{-\frac{1}{2}z^t\Sigma z} \qquad (11)$$

for $|z| \le m^{1/6}$.

Now let $\tilde{f}_m$ and $\tilde{\psi}_m$ be respectively the density and characteristic function of $\mathbb{P}_\epsilon * \theta_\epsilon$. Then $\tilde{\psi}_m(z) = \hat{\theta}_\epsilon(z)\psi_m(z)$. Now we have $|\hat{\theta}_\epsilon(z) - 1| \le C_4\epsilon^{n+1}|z|$ for all z, and so $|\tilde{\psi}_m(z) - \psi_m(z)| \le C_5\epsilon^{n+1}|z|\exp(-\frac{1}{4}z^t\Sigma z)$ for $|z| \le m^{1/2}\delta$. Then (11) holds with ψ replaced by $\tilde{\psi}$ (and a possibly different constant). We also have $|\hat{\theta}_\epsilon(z)| \le C_6\epsilon^{-(q+1)(n+1)}|z|^{-q-1}$ and hence $|\tilde{\psi}_m(z)| \le \gamma^m \min(1, C_6\epsilon^{-(q+1)(n+1)}|z|^{-q-1})$ for $|z| \ge m^{1/2}\delta$. And for $|z| \le m^{1/2}\delta$ we have $|\tilde{\psi}_m(z)| \le \exp(-\frac{1}{4}z^t\Sigma z)$.

Putting all these inequalities together we find that

$$\int_{\mathbb{R}^q} \left| \tilde{\psi}_m(z) - e^{-\frac{1}{2}z^t\Sigma z} \left(1 + \sum_{k=1}^{n-2} m^{-k/2} P_k(z) \right) \right| dz \le C_7 m^{(1-n)/2} \tag{12}$$

and taking inverse Fourier transforms gives

$$\left| \tilde{f}_m(x) - \phi_\Sigma(x) \left(1 + \sum_{k=1}^{n-2} m^{-k/2} Q_k(x) \right) \right| \le C_8 m^{(1-n)/2} \tag{13}$$

for all $x \in \mathbb{R}^q$.

Now, given $r \in N$ and $M > 0$, we apply (13) with $n = r + 3$ and get

$$\left| \tilde{f}_m(x) - \phi_\Sigma(x) \left(1 + \sum_{k=1}^{r+1} m^{-k/2} Q_k(x) \right) \right| \le C_9 m^{-(r+2)/2}$$

Since X_1 has all moments finite, for given $R > 0$ we also have

$$\int_{\mathbb{R}^q} (1 + |x|)^R \left| \tilde{f}_m(x) - \phi_\Sigma(x) \left(1 + \sum_{k=1}^{r+1} m^{-k/2} Q_k(x) \right) \right| dx \le C_{10}$$

and if R was chosen large enough we can deduce from Hölder that

$$\int_{\mathbb{R}^q} (1 + |x|)^M \left| \tilde{f}_m(x) - \phi_\Sigma(x) \left(1 + \sum_{k=1}^{r+1} m^{-k/2} Q_k(x) \right) \right| dx \le C_{11} m^{-(r+1)/2}$$

This inequality still holds (with a different constant) if we remove the $k = r + 1$ term from the sum on the left, and this completes the proof. □

We can now deduce a bound for the CLT in $\mathbb{W}_M$ distance: under the hypotheses of Proposition 2 (and still using $\epsilon = m^{-1/2}$) it follows from this proposition and Theorem 4 that the $\mathbb{W}_M$ distance from $\mathbb{P}_\epsilon$ to $N(0, \Sigma)$ is $O(\epsilon)$. For $q = 1$ this was shown by Rio [8], under different hypotheses, which for $M \le 2$ are significantly less restrictive than ours. We are not however aware of any previous results of this sort in dimension > 1.

By similar arguments we can see that the Edgeworth expansion is an asymptotic expansion of $\mathbb{P}_\epsilon$, in the sense that for any n the $\mathbb{W}_M$ distance from $\mathbb{P}_\epsilon$ to the measure with density $\phi_\Sigma(1 + \sum_{k=1}^n \epsilon^k Q_k)$ is $O(\epsilon^{n+1})$. (This statement is not strictly correct in that this density is not positive everywhere; one should modify it outside a suitable large ball to make it a probability density. Any reasonable way of doing this will give the stated bound).

As with the stochastic Taylor expansion, we can define the sequence of polynomials $(p_1, p_2, \cdots) = \mathcal{S}^{-1}(Q_1, Q_2, \cdots)$. Then Theorem 4 implies that the $\mathbb{W}_M$ distance between $X + \sum_{j=1}^n \epsilon^j p_j(X)$ and $\mathbb{P}_\epsilon$ is $O(\epsilon^{n+1})$. This may be useful for the approximate simulation of Y_m when m is large.

We have assumed here that X has all moments finite; we have not attempted to determine which moments are needed to get $\mathbb{W}_M$ bounds for given M, or to get explicit bounds in terms of the moments. For a detailed investigation of these matters see [12].

5 Asymptotic expansions for $\mathbb{W}_2$ distances

In this section we derive an asymptotic expansion for $\mathbb{W}_2(\mathbb{P}_\epsilon, \tilde{\mathbb{P}}_\epsilon)$ in powers of ϵ, in the situation of Theorem 4. In fact, we consider a more general situation, in that the two $\mathcal{A}$-sequences are allowed to have different covariance matrices. For simulation applications it is probably only the case of equal covariances, and only the leading term of the expansion, that would be of interest. But in the next section we give an application to monotonicity of the $\mathbb{W}_2$ distance which uses the full expansion in the general case.

The derivation of the expansion is based on the following lemma, which is an elaboration of an elementary part of the standard theory of optimal transport for quadratic cost, as described for example in [11].

Lemma 5. *(a) Suppose that ϕ is a C^1 convex real function on $\mathbb{R}^q$ and that X and Z are $\mathbb{R}^q$-valued random variables with the same distribution on the same probability space. Let $Y = g(X)$ where $g = \nabla\phi$. Assume $\mathbb{E}(|X|^2)$, $\mathbb{E}|\phi(X)|$ and $\mathbb{E}(|Y|^2)$ are finite. Then $\mathbb{E}(|Z-Y|^2) \geq \mathbb{E}(|X-Y|^2)$.*
(b) Assume the same hypotheses as in (a), and in addition that ϕ is C^2 and that there exists $\delta > 0$ such that the lowest eigenvalue of $D^2\phi(x)$ is $\geq \delta$ for all x. Then

$$\mathbb{E}|Z-X|^2 \leq \delta^{-1}(\mathbb{E}|Z-Y|^2 - \mathbb{E}|X-Y|^2)$$

Proof. (a) Since ϕ is convex we have $\phi(Z) - \phi(X) \geq (Z-X).Y$ always. Taking expectations, and using $\mathbb{E}\phi(X) = \mathbb{E}\phi(Z)$, then gives $\mathbb{E}(Z.Y) \leq \mathbb{E}(X.Y)$. And since $\mathbb{E}(|X|^2) = \mathbb{E}(|Z|^2)$ we conclude that $\mathbb{E}|Z-Y|^2 \geq \mathbb{E}|X-Y|^2$.
(b) Using the stronger condition on ϕ we have $\phi(z) - \phi(x) - y.(z-x) \geq \frac{\delta}{2}|z-x|^2$ for all x, y, z and then the result follows as in the proof of (a). □

By taking the infimum over all choices of joint distribution of X and Z, we can deduce the following corollary from part (a):

Corollary 6. *Suppose that ϕ is a C^1 convex real function on $\mathbb{R}^q$ and that X is a $\mathbb{R}^q$-valued random variable. Let $Y = g(X)$ where $g = \nabla\phi$. Assume $\mathbb{E}(|X|^2)$, $\mathbb{E}|\phi(X)|$ and $\mathbb{E}(|Y|^2)$ are finite. Then $\mathbb{W}_2(X,Y) = (\mathbb{E}|X-Y|^2)^{1/2}$.*

We mention here a simple application of Corollary 6. Suppose X is an $\mathbb{R}^q$-valued random variable with mean 0 and covariance matrix Σ, and let $Y = AX$ where A is a fixed positive-definite $q \times q$ matrix. Then by applying Corollary 6 with $\phi(x) = \frac{1}{2}x^t Ax$ so that $g(x) = Ax$, we can see that the coupling given by $Y = g(X)$ attains the $\mathbb{W}_2$ distance between X and Y. Hence

$$\mathbb{W}_2(X,Y)^2 = \mathbb{E}|(I-A)X|^2 = \mathrm{tr}((I-A)^2\Sigma) \tag{14}$$

This applies in particular when X and Y have respectively $N(0,\Sigma)$ and $N(0,\tilde{\Sigma})$ distributions, Σ and $\tilde{\Sigma}$ being positive definite. Then there is a unique positive definite A satisfying $A\Sigma A = \tilde{\Sigma}$, given by $A = \Sigma^{-1/2}(\Sigma^{1/2}\tilde{\Sigma}\Sigma^{1/2})^{1/2}\Sigma^{-1/2}$. This implies that AX has the same distribution as Y, so we have $\mathbb{W}_2(X,Y)^2 = \mathrm{tr}((I-A)^2\Sigma)$. For more on this see Corollary 3.2.13 and Theorem 3.4.1 in volume 1 of [7].

Lemma 7. *Suppose $(\mathbb{P}_\epsilon : \epsilon \in E)$ is a family of probability measures on $\mathbb{R}^q$ such that $\sup_{\epsilon\in E}\int |x|^M d\mathbb{P}_\epsilon(x) < \infty$ for each $M > 0$. Let A be a positive definite symmetric $q \times q$ matrix and let $u_1, \cdots, u_k \in P$. For $\epsilon \in E$ define $\rho_\epsilon : \mathbb{R}^q \to \mathbb{R}^q$ by $\rho_\epsilon(x) = Ax + \sum_{j=1}^k \epsilon^j p_j(x)$ where $p_j = \nabla u_j$.*

Then $\mathbb{W}_2(\mathbb{P}_\epsilon, \rho_\epsilon(\mathbb{P}_\epsilon))^2 = \int |x-\rho_\epsilon(x)|^2 d\mathbb{P}_\epsilon(x) + O(\epsilon^M)$ for all $M > 0$.

Proof. The idea is to apply Corollary 6 to a truncated modification $\tilde{\rho}_\epsilon$ of ρ_ϵ which is the gradient of a convex function.

We start by fixing $\beta > 0$ so that β^{-1} exceeds the maximum degree of the polynomials p_j, and then set $R = \epsilon^{-\beta}$. Fix also a smooth function $\psi_0 : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\psi_0(r) = 1$ for $0 < r < 1$ and $\psi_0(r) = 0$ for $r > 2$. Then define $\psi(r) = \psi_0(r/R)$ for $r \in \mathbb{R}_+$. Next define

$\tilde{u}_j(x) = \psi(|x|)u_j(x)$, $\tilde{u}_\epsilon(x) = \frac{1}{2}x^tAx + \epsilon^j\tilde{u}_j(x)$ and $\tilde{\rho}_\epsilon(x) = \nabla\tilde{u}_\epsilon(x) = Ax + \sum_{j=1}^k \epsilon^j\tilde{p}_j(x)$ for $x \in \mathbb{R}^q$, where $\tilde{p}_j = \nabla\tilde{u}_j$.

Our choice of R ensures that for ϵ small enough $\tilde{u}_\epsilon$ will be convex and then by Corollary 6 we have

$$\mathbb{W}_2(\mathbb{P}_\epsilon, \tilde{\rho}_\epsilon(\mathbb{P}_\epsilon))^2 = \int |x - \tilde{\rho}(x)|^2 d\mathbb{P}_\epsilon(x) \tag{15}$$

And we see that

$$\rho_\epsilon(x) - \tilde{\rho}_\epsilon(x) = (1 - \psi(|x|))\rho_\epsilon(x) - \frac{\psi'(|x|)}{|x|}\sum_{j=1}^k \epsilon^j u_j(x)x \tag{16}$$

which vanishes if $|x| \leq R$. Then we have

$$\mathbb{W}_2(\rho_\epsilon(\mathbb{P}_\epsilon), \tilde{\rho}_\epsilon(\mathbb{P}_\epsilon))^2 \leq \int |\rho_\epsilon(x) - \tilde{\rho}_\epsilon(x)|^2 d\mathbb{P}_\epsilon(x) = O(\epsilon^M) \tag{17}$$

for any $M > 0$. Combining (15) and (17), and using the triangle inequality, gives the result. □

We next introduce a generalisation of the map $\mathcal{S}_\Sigma$. Let Σ and A be positive definite $q \times q$ matrices, and let $\tilde{\Sigma} = A\Sigma A$. Then we define $\mathcal{S}_{\Sigma,A} : \mathcal{P}_\Sigma \times \mathcal{P}_G^q \to \mathcal{P}_{\tilde{\Sigma}}$ by formal power series operations as follows: let $\mathbf{S} = (S_1, S_2, \cdots) \in \mathcal{P}_\Sigma$ and let $\mathbf{p} = (p_1, p_2, \cdots) \in \mathcal{P}_G^q$. Then we determine formally the image of the measure with density $(1 + \epsilon S_1 + \epsilon^2 S_2 + \cdots)\phi_\Sigma$ under the mapping $y = \psi_\epsilon(x) := Ax + \epsilon p_1(x) + \epsilon^2 p_2(x) + \cdots$, obtaining a formal expansion

$$\det\{D\psi_\epsilon^{-1}(y)\}\phi_\Sigma(\psi_\epsilon^{-1}(y)) = (1 + \epsilon\tilde{S}_1(y) + \epsilon^2\tilde{S}_2(y) + \cdots)\phi_{\tilde{\Sigma}}(y)$$

for the density of the image, where $\tilde{S}_j \in P_{\tilde{\Sigma}}$. Then $\mathcal{S}_{\Sigma,A}(\mathbf{S}, \mathbf{p}) = \tilde{\mathbf{S}} := (\tilde{S}_1, \tilde{S}_2, \cdots)$. We note that in the special case where $\mathbf{p} = (0, 0, 0, \cdots)$, so that $\psi_\epsilon(x) = Ax$, we get $\tilde{S}_k(y) = S_k(A^{-1}y)$.

The following observation about the parity of such expansion will be used. We say that $\mathbf{p} \in \mathcal{P}$ or $\mathcal{P}^q$ is *even* if $p_k(-x) = (-1)^k p(x)$ and *odd* if $p_k(-x) = (-1)^{k+1}p(x)$ for all k and x. Then if $\mathbf{S} \in \mathcal{P}_\Sigma$ is even and $\mathbf{p} \in \mathcal{P}^q$ is odd, we easily see that $\mathcal{S}_{\Sigma,A}(\mathbf{S}, \mathbf{p})$ is even.

We aim to obtain an estimate for Vaserstein 2-distances in the context of the last paragraph. To this end we consider another way of (formally) describing the mapping ψ_ϵ, by first finding $(q_1, q_2, \cdots)$ so that $\mathcal{S}_\Sigma(q_1, q_2, \cdots) = (S_1, S_2, \cdots)$ and then writing the formal composition of the mapping $x = A^{-1}v + \epsilon q_1(A^{-1}v) + \cdots$ with $y = \psi_\epsilon(x)$ as $y = v + \epsilon r_1(v) + \cdots$. Then we have:

Lemma 8. *With notation as above,* $\mathcal{S}_{\tilde{\Sigma}}(r_1, r_2, \cdots) = (\tilde{S}_1, \tilde{S}_2, \cdots)$.

Proof. This is fairly clear from consideration of the formal power series, but can also be seen as follows: fix $n \in \mathbb{N}$ and a ball $B = \{|x| \leq R\}$ in $\mathbb{R}^q$. Then for ϵ small enough, the image under $y = Ax + \sum_{k=1}^n \epsilon^k p_k(x)$ of the restriction to B of the density $(1 + \sum_{k=1}^n \epsilon^k S_k)\phi_\Sigma$ will be a density $(1 + \sum_{k=1}^n \epsilon^k \tilde{S}_k)\phi_{\tilde{\Sigma}} + O(\epsilon^{n+1})$. On the other hand this density also agrees, up to $O(\epsilon^{n+1})$, by the image under $y = v + \sum_{k=1}^n \epsilon^k r_k(v)$, and the result follows. □

Note that one consequence of Lemma 8, along with Lemma 1, is that $\tilde{S}_j \in P_{\tilde{\Sigma}}$. We can now state the desired estimate:

Proposition 3. *Suppose* $\Sigma, A, \tilde{\Sigma}, \mathbf{p}, \mathbf{S}, \tilde{\mathbf{S}}$ *are as above, with* $\tilde{\mathbf{S}} = \mathcal{S}_{\Sigma,A}(\mathbf{S}, \mathbf{p})$. *Suppose also that* $(\mathbb{P}_\epsilon : \epsilon \in E)$ *and* $(\tilde{\mathbb{P}}_\epsilon : \epsilon \in E)$ *are families of probability distributions on* $\mathbb{R}^q$ *with respectively a* $\mathcal{A}_\Sigma$*-sequence* $\mathbf{S}$ *and a* $\mathcal{A}_{\tilde{\Sigma}}$*-sequence* $\tilde{\mathbf{S}}$. *Then for any* $n \in \mathbb{N}$ *we have*

$$\mathbb{W}_2(\mathbb{P}_\epsilon, \tilde{\mathbb{P}}_\epsilon) = \left\{\int_{\mathbb{R}^q} \left|Ax - x + \sum_{k=1}^n \epsilon^k p_k(x)\right|^2 \left(1 + \sum_{k=1}^n \epsilon^k S_k(x)\right)\phi_\Sigma(x)dx\right\}^{1/2} + O(\epsilon^{n+1}) \tag{18}$$

Proof. Let $q_1, q_2, \cdots$ and $r_1, r_2, \cdots$ be as defined in the paragraph before Lemma 8. Let Z be a random variable with $N(0, \Sigma)$ distribution, and define $V = AZ$ and $X_\epsilon = Z + \sum_{k=1}^n \epsilon^n q_k(Z)$. Also set $Y_\epsilon = AX_\epsilon + \sum_{k=1}^n \epsilon^k p_k(X_\epsilon)$, which can be rewritten as

$$Y_\epsilon = V + \sum_{k=1}^n \epsilon^k r_k(V) + \epsilon^{n+1} s(\epsilon, V)$$

for some polynomial s. Then (X_ϵ) has an $\mathcal{A}_\Omega$-sequence agreeing with $\mathbf{S}$ up to order ϵ^n, and so $\mathbb{W}_2(\mathbb{P}_\epsilon, X_\epsilon) = O(\epsilon^{n+1})$ by Theorem 4. Similarly (Y_ϵ) has an $\mathcal{A}_{\tilde{\Omega}}$-sequence agreeing with $\tilde{\mathbf{S}}$ up to order ϵ^n (by Lemma 8), and so $\mathbb{W}_2(\tilde{\mathbb{P}}_\epsilon, Y_\epsilon) = O(\epsilon^{n+1})$. And Lemma 7 gives $\mathbb{W}(X_\epsilon, Y_\epsilon) = (\mathbb{E}|X_\epsilon - Y_\epsilon|^2)^{1/2} + O(\epsilon^{n+1})$. Putting these results together gives (18). □

We remark that the conclusion of Proposition 3 can be reformulated as a statement that $\mathbb{W}_2(\mathbb{P}_\epsilon, \tilde{\mathbb{P}}_\epsilon)^2$ has an asymptotic expansion $\sum_{k=0}^\infty C_k \epsilon^k$, where from (18), and writing $p_0(x) = Ax - x$ and $S_0(x) = 1$, we have

$$C_k = \sum_{i+j+l=k;\ i,j,k\geq 0} \int_{\mathbb{R}^q} p_i(x)^t p_j(x) S_l(x) \phi_\Sigma(x) dx \tag{19}$$

In particular $C_0 = \int_{\mathbb{R}^q} |(A-I)x|^2 \phi_\Sigma(x) dx = \text{tr}((A-I)^2\Sigma)$.

We also remark that, in the special case where $\mathbb{P}_\epsilon$ and $\tilde{\mathbb{P}}_\epsilon$ have covariance matrices independent of ϵ, i.e.

$$\int xx^t d\mathbb{P}_\epsilon(x) = \Sigma, \qquad \int xx^t d\tilde{\mathbb{P}}_\epsilon(x) = \tilde{\Sigma} \tag{20}$$

for each ϵ, we can simplify (19) somewhat by a modified derivation of the asymptotic expansion, as follows. Using the notation of the proof of Proposition 3, for given n we can write $Y_\epsilon = AX_\epsilon + V_\epsilon$ where $V_\epsilon = \sum_{k=1}^n \epsilon^k p_k(X_\epsilon)$ and then from (20) we have $\mathbb{E}(X_\epsilon X_\epsilon^t) = \Sigma + O(\epsilon^{n+1})$ and $\mathbb{E}(Y_\epsilon Y_\epsilon^t) = \tilde{\Sigma} + O(\epsilon^{n+1})$. Now we have

$$\mathbb{E}(Y_\epsilon Y_\epsilon^t) = \mathbb{E}(AX_\epsilon X_\epsilon^t A) + AB_\epsilon + B_\epsilon^t A + G_\epsilon$$

where $B_\epsilon = \mathbb{E}(X_\epsilon V_\epsilon^t)$ and $G_\epsilon = \mathbb{E}(V_\epsilon V_\epsilon^t)$. It follows that $AB_\epsilon + B_\epsilon^t A = -G_\epsilon + O(\epsilon^{n+1})$. Using this, we find that

$$\begin{aligned}\mathbb{E}|Y_\epsilon - X_\epsilon|^2 = \mathbb{E}|(A-I)X_\epsilon + V_\epsilon|^2 &= \text{tr}((A-I)^2\Sigma) + 2\text{tr}((A-I)B_\epsilon) + \text{tr}G_\epsilon + O(\epsilon^{n+1}) \\ &= C_0 + \text{tr}(A^{-1}G_\epsilon) + O(\epsilon^{n+1})\end{aligned}$$

Then using the definition of V_ϵ to expand $G_\epsilon = \mathbb{E}(V_\epsilon V_\epsilon^t)$ we conclude that when (20) holds we have

$$C_k = \sum_{i+j+l=k;\ i>0,j>0,l\geq 0} \int_{\mathbb{R}^q} p_i(x)^t A^{-1} p_j(x) S_l(x) \phi_\Sigma(x) dx \tag{21}$$

for $k > 0$. This is simpler than (19) in that there are no terms involving p_0. In particular we have $C_1 = 0$.

In order to apply apply Proposition 3 to an arbitrary pair of families $(\mathbb{P}_\epsilon)$ and $(\tilde{\mathbb{P}}_\epsilon)$ with $\mathcal{A}$-sequences $\mathbf{S}$ and $\tilde{\mathbf{S}}$ we need to find positive definite A and $\mathbf{p} \in \mathcal{P}_\mathbf{G}^\mathbf{q}$ such that $\tilde{\Sigma} = A\Sigma A$ and $\tilde{\mathbf{S}} = \mathcal{S}_{\Sigma,A}(\mathbf{S}, \mathbf{p})$, and we turn to this now.

As we saw above, there is a unique positive definite A satisfying $A\Sigma A = \tilde{\Sigma}$, given by $A = \Sigma^{-1/2}(\Sigma^{1/2}\tilde{\Sigma}\Sigma^{1/2})^{1/2}\Sigma^{-1/2}$. For the construction of $\mathbf{p}$ we need a couple of lemmas:

Lemma 9. *Let $B = (b_{ij})$ be a diagonalisable $q \times q$ real matrix with positive eigenvalues, let $n \in \mathbb{N}$ and let V_n be the real vector space of homogeneous degree-n polynomials on $\mathbb{R}^q$. Define $T : V_n \to V_n$ by $Tf(x) = \sum_{i,j=1}^q x_j b_{ij} \frac{\partial f}{\partial x_i}$.*

Then T is surjective.

Proof. Let J be an invertible matrix such that $J^{-1}BJ$ =diag$(\lambda_1, \cdots, \lambda_q$ where all $\lambda_j > 0$. Define an invertible linear map $U : V_n \to V_n$ by $Uf(z) = f(Jz)$ for $z \in \mathbb{R}^q$, and then define a linear map $S : V_n \to V_n$ by $S = U^{-1}TU$. Then $Sg(z) = \sum_{i=1}^q \lambda_i z_i \frac{\partial g}{\partial z_i}$ for $g \in V$. So if $g(z) = z_1^{r_1} \cdots z_q^{r_q}$ where $r_1 + \cdots r_q = n$ we have $Sg = \lambda g$ where $\lambda = \sum_{i=1}^q r_i\lambda_i > 0$. Since such monomials form a basis of V_n it follows that S, and hence T, is invertible. □

Lemma 10. *The map $\mathcal{L}_{\Sigma,A}$ defined by*

$$\mathcal{L}_{\Sigma,A}\psi(y) = y^t\tilde{\Sigma}^{-1}\psi(A^{-1}y) - \mathrm{tr}(A^{-1}D\psi(A^{-1}y))$$

maps P_G^q bijectively onto $P_{\tilde{\Sigma}}$.

Proof. First we note that $\phi_{\tilde{\Sigma}}(y)\mathcal{L}_{\Sigma,A}\psi(y)$ is the divergence of $\phi_{\tilde{\Sigma}}(y)\psi(A^{-1}y)$ and hence integrates to 0. So the range of $\mathcal{L}_{\Sigma,A}$ is in $P_{\tilde{\Sigma}}$.

Next, applying Lemma 9 with $B = A\tilde{\Sigma}^{-1}$, we see that if g is a homogeneous polynomial of positive degree on $\mathbb{R}^q$, then there is a unique $\psi \in P_G^q$ such that $x^tB\psi(x) = g(Ax)$, which means (putting $y = Ax$) that $y^t\tilde{\Sigma}^{-1}\psi(A^{-1}y) = g(y)$ for all $y \in \mathbb{R}^q$. From this one can easily deduce, by induction on the degree of f, that for any $f \in P_{\tilde{\Sigma}}$ there is a unique $\psi \in P_G^q$ such that $\mathcal{L}_{\Sigma,A}\psi = f$. □

Proposition 4. *Given $\mathbf{S} \in \mathcal{P}_{\mathbf{\Sigma}}$ and $\tilde{\mathbf{S}} \in \mathcal{P}_{\tilde{\Sigma}}$, there is a unique $\mathbf{p} \in \mathcal{P}_{\mathbf{G}}^{\mathbf{q}}$ such that $\mathcal{S}_{\Sigma,A}(\mathbf{S}, \mathbf{p}) = \tilde{\mathbf{S}}$.*

Proof. We follow the method of proof of Lemma 2. We show by induction on n that for each n there is a unique choice of $p_1, \cdots, p_n \in P_G^q$ such that $\mathcal{S}_{\Sigma,A}^{(n)}(S_1, \cdots, S_n, p_1, \cdots, p_n) = (\tilde{S}_1, \cdots, \tilde{S}_n)$.

So we suppose we have such $p_1, \cdots, p_n$, and look for $p_{n+1} \in P_G^q$ such that

$$\mathcal{S}_{\Sigma,A}^{(n+1)}(S_1, \cdots, S_{n+1}, p_1, \cdots, p_{n+1}) = (\tilde{S}_1, \cdots, \tilde{S}_{n+1})$$

We can write $\mathcal{S}_{\Sigma,A}^{(n+1)}(S_1, \cdots, S_{n+1}, p_1, \cdots, p_n) = (\tilde{S}_1, \cdots, \tilde{S}_n, v)$ for some $v \in P_{\tilde{\Sigma}}$. Then for any choice of $p_{n+1} \in P_G^q$, we have

$$\mathcal{S}_{\Sigma,A}^{(n+1)}(S_1, \cdots, S_{n+1}, p_1, \cdots, p_{n+1}) = (\tilde{S}_1, \cdots, \tilde{S}_n, v - \mathcal{L}_{\Sigma,A}p_{n+1})$$

So we need p_{n+1} to satisfy $\mathcal{L}_{\Sigma,A}p_{n+1} = v - \tilde{S}_{n+1}$, and by Lemma 10 there is a unique such $p_{n+1} \in P_G^q$. This completes the inductive step. The initial step is proved in the same way. □

Combining Propositions 3 and 4, we obtain the main result of this section:

Theorem 11. *Suppose that $(\mathbb{P}_\epsilon : \epsilon \in E)$ and $(\tilde{\mathbb{P}}_\epsilon : \epsilon \in E)$ are families of probability distributions on $\mathbb{R}^q$ with an $\mathcal{A}_\Sigma$-sequence $\mathbf{S}$ and an $\mathcal{A}_{\tilde{\Sigma}}$-sequence $\tilde{\mathbf{S}}$ respectively, where Σ and $\tilde{\Sigma}$ are positive definite. Then $\mathbb{W}_2(\mathbb{P}_\epsilon, \tilde{\mathbb{P}}_\epsilon)^2$ has an asymptotic expansion $\sum_{k=0}^\infty C_k\epsilon^k$, in the sense that for each n there exists $K > 0$ such that*

$$\left| \mathbb{W}_2(\mathbb{P}_\epsilon, \tilde{\mathbb{P}}_\epsilon)^2 - \sum_{k=0}^n C_k\epsilon^k \right| \le K\epsilon^{n+1}$$

for all $\epsilon \in E$.

Here C_k is given by (19), where $\mathbf{p} = (p_1, p_2, \cdots)$ is given by Proposition 4 and $p_0(x) = Ax - x$, $S_0(x) = 1$.

We remark that, from (19) and the construction of the p_j, it is not hard to see that C_k can be expressed as a polynomial in the entries of Σ and A, $\det(\Sigma)^{-1}$, and the coefficients of $S_1, \cdots, S_k$ and $\tilde{S}_1, \cdots, \tilde{S}_k$.

We also remark that, if $\mathbf{S}$ and $\tilde{\mathbf{S}}$ are both even, then using the earlier observation on parity and the construction in the proof of Proposition 4, we obtain by induction that $\mathbf{p}$ is odd, and then it follows by (19) that $C_k = 0$ if k is odd.

A further remark is that, when (20) holds, when can use (21) to simplify the calculation of C_k. We have $C_1 = 0$, and we also get a simple expression for the first non-zero C_k (not counting C_0), as follows. Excluding the case where all p_m vanish, let m be the smallest number in $\mathbb{N}$ such that p_m is not identically zero. Equivalently m is the smallest number such that $\tilde{S}_m(x) \neq S_m(A^{-1}x)$ for some x. Then by (21) we have $C_k = 0$ for $1 \leq k < 2m$ while $C_{2m} = \int_{\mathbb{R}^q} p_m(x)^t A^{-1} p_m(x) \phi_\Sigma(x) dx$. It follows from this and the fact that p_m does not vanish identically that $C_{2m} > 0$. Also we see from the construction in the proof of Proposition 4 that p_m is the unique polynomial in $\mathcal{P}_G^q$ satisfying $\mathcal{L}_{\Sigma,A} p_m = S_m(A^{-1}x) - \tilde{S}_m$.

We conclude this section with a result which can be regarded as an asymptotic expansion for an optimal coupling (for the quadratic distance) between $\mathbb{P}_\epsilon$ and $\tilde{\mathbb{P}}_\epsilon$. We first note that it is elementary that optimal couplings exist (see e.g. Proposition 2.1 of [11]). Then we have following result for such couplings.

Proposition 5. *Let Σ, A, $\tilde{\Sigma}$, $(\mathbb{P}_\epsilon)$ and $(\tilde{\mathbb{P}}_\epsilon)$ be as above. and let $k \in \mathbb{N}$. Then there is $C > 0$ such that, if Y and Z are random vectors on the same probability space, with distributions $(\mathbb{P}_\epsilon)$ and $(\tilde{\mathbb{P}}_\epsilon)$ respectively, such that $\mathbb{E}|Z - Y|^2 = \mathbb{W}_2(\mathbb{P}_\epsilon, \tilde{\mathbb{P}}_\epsilon)^2$, then*

$$\mathbb{E}\left|Z - AY - \sum_{j=1}^{k} \epsilon^j p_j(Y)\right|^2 \leq C\epsilon^{2k+2} \tag{22}$$

Proof. Fix $k \in \mathbb{N}$ and let $\rho_\epsilon(y) = Ay + \sum_{j=1}^{2k+1} \epsilon^j p_j(y)$. Now define $\tilde{\rho}_\epsilon = \nabla \tilde{u}_\epsilon$ in the same way as in the proof of Lemma 7; then for ϵ small enough, $\tilde{u}_\epsilon$ is strictly convex and smooth, and $\tilde{\rho}_\epsilon$ maps $\mathbb{R}^q \to \mathbb{R}^q$ bijectively with inverse $g_\epsilon = \nabla \phi_\epsilon$ where ϕ_E is convex and smooth. Also, for ϵ small enough, $\|D\tilde{\rho}_\epsilon\| \leq 2\|A\|$ everywhere, so $D^2\phi_\epsilon = Dg_\epsilon$ has smallest eigenvalue $\geq (2\|A\|)^{-1}$.

Now let Y and Z be as in the statement, and let $X = \tilde{\rho}_\epsilon(Y)$, so that $Y = g_\epsilon(X)$. Then Lemma 5(b) applies with $\phi = \phi_\epsilon$ and, recalling that $\mathbb{E}|Z - Y|^2 = \mathbb{W}_2(\mathbb{P}_\epsilon, \tilde{\mathbb{P}}_\epsilon)^2$, we see that

$$\mathbb{E}|Z - X|^2 \leq 2\|A\|(\mathbb{W}_2(\mathbb{P}_\epsilon, \tilde{\mathbb{P}}_\epsilon)^2 - \mathbb{E}|X - Y|^2)$$

Now arguing as in the proof of Proposition 3 we find that $\mathbb{W}_2(\mathbb{P}_\epsilon, \tilde{\mathbb{P}}_\epsilon)^2 - \mathbb{E}|X - Y|^2 \leq C\epsilon^{2k+2}$, so that $\mathbb{E}|Z - X|^2 = O(\epsilon^{2k+2})$. Also, as in the proof of Lemma 7 we have $\mathbb{E}|X - \rho_\epsilon(Y)|^2 = O(\epsilon^{2k+2})$. Moreover $\mathbb{E}|\sum_{j=k+1}^{2k+1} \epsilon^j p_j(Y)|^2 = O(\epsilon^{2k+2})$, and the result follows from these last three bounds. □

We remark that in general optimal couplings may not be unique, but if $\mathbb{P}_\epsilon$ has the property that $\mathbb{P}_\epsilon(F) = 0$ for any Borel set F of Hausdorff dimension $q - 1$, then there is a unique optimal coupling given by $Z = \psi_\epsilon(Y)$, where $\psi_\epsilon : \mathbb{R}^q \to \mathbb{R}^q$ is the gradient of a convex function (see Theorem 2.12 of [11]). In this case we can rewrite the conclusion (22) as

$$\int_{\mathbb{R}^q} \left|\psi_\epsilon(y) - Ay - \sum_{j=1}^{k} \epsilon^j p_j(y)\right|^2 d\mathbb{P}_\epsilon(y) \leq C\epsilon^{2k+2}$$

In this sense the series $Ay + \sum_{j=1}^{\infty} \epsilon^j p_j(y)$ can be regarded as an asymptotic expansion of ψ_ϵ.

We also note that if $\mathbb{P}_\epsilon$ and $\tilde{\mathbb{P}}_\epsilon$ have densities f_ϵ and $\tilde{f}_\epsilon$ respectively, then under suitable regularity conditions, ψ_ϵ satisfies a Monge-Ampere-type equation $\det(D\psi_\epsilon(y))\tilde{f}_\epsilon(\psi_\epsilon(y)) =$

$f_\epsilon(y)$. See Chapter 4 of [11] for detailed discussion of this. We merely remark here that the series $Ay + \epsilon p_1(y) + \epsilon^2 p_2(y) + \cdots$ satisfies the above Monge-Ampere equation in a formal power series sense (using the the $\mathcal{A}$-sequences as formal series for f_ϵ and $\tilde{f}_\epsilon$).

6 Asymptotic expansions in Central Limit case

Now we apply Theorem 11 to obtain asymptotic expansions for central limit $\mathbb{W}_2$ distances using Proposition 2. We suppose X and $\tilde{X}$ are $\mathbb{R}^q$-valued random variables, with zero mean and covariance matrices Σ and $\tilde{\Sigma}$ respectively. We also assume that X and $\tilde{X}$ have all moments finite, and satisfy the CC condition (which implies that Σ and $\tilde{\Sigma}$ are invertible). Let $Y_m = m^{-1/2}\sum_{k=1}^m X_k$ and $\tilde{Y}_m$ similarly. Next let $\mathbf{Q} = (Q_1, Q_2, \cdots)$ be the $\mathcal{A}$-sequence for the distributions of the sequence (Y_m) given by Proposition 2, and likewise $\tilde{\mathbf{Q}}$.

Then we apply Theorem 11 with $E = \{m^{-1/2} : m \in \mathbb{N}\}$ and with $\mathbb{P}_\epsilon$ and $\tilde{\mathbb{P}}_\epsilon$ being the distributions of Y_m and $\tilde{Y}_m$ respectively, where $\epsilon = m^{-1/2}$. We also note that $\mathbf{Q}$ and $\tilde{\mathbf{Q}}$ are both even and that (20) holds, so that by the remarks at the end of the last section we have an asymptotic expansion $\sum_{k=0}^\infty C_k m^{-k/2}$ for $\mathbb{W}_2(Y_m, \tilde{Y}_m)^2$, where C_k for $k > 0$ is given by (21) with $\mathbf{S} = \mathbf{Q}$ and $\tilde{\mathbf{S}} = \tilde{\mathbf{Q}}$, and $C_k = 0$ for odd k. Then, writing $B_k = C_{2k}$, we can express the asymptotic expansion as $\sum_{k=0}^\infty B_k m^{-k}$ where $B_0 = \mathrm{tr}((A-I)^2\Sigma)$ and for $k > 0$

$$B_k = \sum_{i+j+l=2k; i>0, j>0, l\geq 0} \int_{\mathbb{R}^q} p_i(x)^t A^{-1} p_j(x) Q_l(x) \phi_\Sigma(x) dx \tag{23}$$

where $\mathbf{p} \in \mathcal{P}_G^q$ satisfies $\tilde{\mathbf{Q}} = \mathcal{S}_{\Sigma,A}(\mathbf{Q}, \mathbf{p})$ and as before $Q_0(x) = 1$.

We remark that in the case where X and $\tilde{X}$ both have symmetric distributions (i.e. X has the same distribution as $-X$, and likewise for $\tilde{X}$) then we have $Q_j = 0$ and $\tilde{Q}_j = 0$ for j odd, from which it is easy to deduce that $p_j = 0$ for j odd. This simplifies the calculation of B_k in (23). In particular we have $B_1 = 0$ in this case.

In general the calculation of the coefficients B_k using (23) is quite heavy; we now give two calculations for relatively simple cases. The first finds B_1 for the usual 'CLT' situation where one of the distributions is normal (with the same covariance), while the second finds B_1 and B_2 when $q = 1$.

Calculation of B_1 when one distribution is normal

We suppose X has mean 0 and covariance Σ (and as usual that X has all moments finite and satisfies the CC condition), and we seek an asymptotic expansion $\mathbb{W}_2(X, N(0,\Sigma))^2 \sim \sum B_k m^{-k}$. By an orthogonal change of coordinates, which does not change the Vaserstein distance, we can suppose Σ is diagonal with entries $\sigma_1^2, \cdots, \sigma_q^2$ where each $\sigma_j > 0$. We also write $\mu_{jkl} = \mathbb{E}(X_j X_k X_l)$ for the 3rd moments, where $j, k, l = 1, 2, \cdots, q$. We recall the eigenfunctions of $\mathcal{S}_\Sigma \nabla$ described in section 2. Then Q_1 can be expanded in terms of these eigenfunctions:

$$Q_1(x) = \sum_{j<k<l} \mu_{jkl} \frac{x_j x_k x_l}{\sigma_j^2 \sigma_k^2 \sigma_l^2} + \frac{1}{2}\sum_{j\neq k} \mu_{jjk} \frac{H_2(x_j/\sigma_j) x_k}{\sigma_j^2 \sigma_k^2} + \frac{1}{6}\sum_j \mu_{jjj} \frac{H_3(x_j/\sigma_j)}{\sigma_j^3}$$

Now we need $p_1 = \nabla u$ where u satisfies $\mathcal{S}_\Sigma \nabla u = Q_1$. Using the eigenvalues of $\mathcal{S}_\Sigma \nabla$ given in section 2 we find

$$u(x) = \sum_{j<k<l} \mu_{jkl} \frac{x_j x_k x_l}{\sigma_k^2\sigma_l^2 + \sigma_j^2\sigma_l^2 + \sigma_j^2\sigma_k^2} + \frac{1}{2}\sum_{j\neq k} \mu_{jjk} \frac{H_2(x_j/\sigma_j) x_k}{\sigma_j^2 + 2\sigma_k^2} + \frac{1}{18}\sum_j \mu_{jjj} \frac{H_3(x_j/\sigma_j)}{\sigma_j}$$

Then we calculate $p_1 = \nabla u$ and find

$$B_1 = \int p_1(x)^2 \phi_\Sigma(x) dx = \sum_{j<k<l} \frac{\mu_{jkl}^2}{\sigma_k^2\sigma_l^2 + \sigma_j^2\sigma_l^2 + \sigma_j^2\sigma_k^2} + \frac{1}{2}\sum_{j\neq k} \frac{\mu_{jjk}^2}{\sigma_j^2(\sigma_j^2 + 2\sigma_k^2)} + \frac{1}{18}\sum_j \frac{\mu_{jjj}^2}{\sigma_j^4}$$

which can be rearranged as

$$B_1 = \frac{1}{6} \sum_{j,k,l=1}^{q} \frac{\mu_{jkl}^2}{\sigma_k^2\sigma_l^2 + \sigma_j^2\sigma_l^2 + \sigma_j^2\sigma_k^2}$$

So we can deduce that

$$\mathbb{W}_2(Y_m, N(0,\Sigma))^2 = \frac{1}{6} \sum_{j,k,l=1}^{q} \frac{\mu_{jkl}^2}{\sigma_k^2\sigma_l^2 + \sigma_j^2\sigma_l^2 + \sigma_j^2\sigma_k^2} m^{-1} + O(m^{-2}) \tag{24}$$

One-dimensional case

We first note that when $q = 1$ we can always reduce to the case of equal variance 1, because of the observation that if U and V are random variables with mean 0 and variance 1, and if λ and ρ are positive constants, then for any joint distribution of U, V we have $\mathbb{E}(\lambda U - \rho V)^2 = (\lambda - \rho)^2 + \lambda\rho\mathbb{E}(U - V)^2$, from which we deduce that $\mathbb{W}_2(\lambda U, \rho V)^2 = (\lambda - \rho)^2 + \lambda\rho\mathbb{W}_2(U, V)^2$.

So we suppose $q = 1$ and that X and $\tilde{X}$ are random variables with mean 0 and variance 1, with all moments finite and satisfying the CC condition as above. For $k \in \mathbb{N}$ define $\mu_k = \mathbb{E}(X^k)$ and $\tilde{\mu}_k = \mathbb{E}(\tilde{X}^k)$. We note that in this case $\mathcal{L}_\Sigma$ reduces to $\mathcal{L}_1 p(x) = xp(x) - p'(x)$, and again it will be useful to expand polynomials in terms of Hermite polynomials, and use the relation $\mathcal{L}_1 H_k = H_{k+1}$. Now for the first term in the Edgeworth expansions we find $Q_1 = \frac{1}{6}\mu_3 H_3$ and similarly for $\tilde{Q}_1$. Then we have $A = 1$ and we calculate $p_1 = \frac{1}{6}(\tilde{\mu}_3 - \mu_3)H_2$. Finally we obtain $B_0 = 0$ and, using (23), we calculate $B_1 = \int_{\mathbb{R}} p_1(x)^2\phi(x)dx = \frac{1}{18}(\tilde{\mu}_3 - \mu_3)^2$. We conclude that

$$\mathbb{W}_2(Y_m, \tilde{Y}_m)^2 = \frac{1}{18}(\tilde{\mu}_3 - \mu_3)^2 m^{-1} + O(m^{-2}) \tag{25}$$

In the usual CLT situation, $\tilde{X}$ has $N(0,1)$ distribution so $\tilde{\mu}_3 = 0$. Then (25) becomes

$$\mathbb{W}_2(Y_m, N(0,1))^2 = \frac{\mu_3^2}{18m} + O(m^{-2}) \tag{26}$$

which is a special case of (24). Now we compare (26) with the results of Rio [9]. Theorem 1.1 of [9] states that, if $1 < r \le 2$ and X is a random variable with mean 0, variance 1 and $\mathbb{E}|X|^{r+2} < \infty$, whose distribution is not supported on a lattice, then

$$\mathbb{W}_r(Y_m, N(0,1)) = \frac{\lambda_r|\mu_3|}{6} m^{-1/2} + o(m^{-1/2}) \tag{27}$$

where $\lambda_r = (\int_{\mathbb{R}} |1 - x^2|^r \phi(x)dx)^{1/r}$. We have $\lambda_2 = \sqrt{2}$ so (27) with $r = 2$ is consistent with (26). In fact (27) with $r = 2$ is equivalent to (26) with $o(m^{-1})$ in place of $O(m^{-2})$, so (26) gives a stronger bound - but it also requires a stronger moment condition, as well as the CC condition which is stronger than the non-lattice condition in Theorem 1.1 of [9].

We also note that Rio gives an analogous result (Theorem 1.2 of [9]) for lattice distributions. It would be of interest to investigate to what extent the results of this section could be extended to lattice distributions and also to $\mathbb{W}_r$ bounds for $r \ne 2$.

We can extend the expansion in (25) to higher order using (23), though the complexity of the calculation of B_k increases rapidly with k. To calculate B_2, we note first that (23) gives

$$B_2 = \int_{\mathbb{R}} \{2p_1(x)p_3(x) + p_2(x)^2 + 2p_1(x)p_2(x)Q_1(x) + p_1(x)^2 Q_2(x)\}\phi(x)dx \tag{28}$$

Now from the Edgeworth expansion again we have $Q_2(x) = \frac{1}{24}(\mu_4 - 3)H_4(x) + \frac{1}{72}\mu_3^2 H_6(x)$, and we also find after some calculation that $p_2(x) = JH_3(x) - Kx$ where $J = \frac{1}{24}(\tilde{\mu}_4 - \mu_4) -$

$\frac{1}{18}(\tilde{\mu}_3 - \mu_3)(\tilde{\mu}_3 + 2\mu_3)$ and $K = \frac{1}{36}(\tilde{\mu}_3 - \mu_3)(\tilde{\mu}_3 + 5\mu_3)$. We also need p_3, which contains a large number of terms so we omit the details, but remark that the calculation is simplified by the observation that, since p_1 is a constant multiple of H_2, to determine $\int p_1 p_3 \phi$ we only need the coefficient of H_2 in the expansion of p_3 in Hermite polynomials. When this is found and everything is put together we obtain

$$\begin{aligned} B_2 =& \frac{1}{96}(\mu_4 - \tilde{\mu}_4)^2 + \frac{1}{36}(\mu_3 - \tilde{\mu}_3)(-2\mu_3\mu_4 + 2\tilde{\mu}_3\tilde{\mu}_4 + \mu_3\tilde{\mu}_4 - \tilde{\mu}_3\mu_4) \\ &+ \frac{1}{12}(\mu_3 - \tilde{\mu}_3)^2 \left(1 + \frac{1}{324}(227\mu_3^2 + 302\mu_3\tilde{\mu}_3 + 227\tilde{\mu}_3^2)\right) \end{aligned} \tag{29}$$

In the situation where $\tilde{X}$ is normally distributed, we have $\tilde{\mu}_3 = 0$ and $\tilde{\mu}_4 = 3$, and then as noted above $B_1 = \frac{\mu_3^2}{18}$, while (29) reduces to

$$B_2 = \frac{1}{96}(\mu_4 - 3)^2 - \frac{1}{18}\mu_3^2(\mu_4 - 3) - \frac{227}{3888}\mu_3^4 \tag{30}$$

In this case the expansion $x + \sum m^{-k/2} p_k(x)$ is a Cornish-Fisher series, and we can use the known expressions for such series to give a simpler derivation of (30). But we can also go further and use Cornish-Fisher expansions to simplify he proof of (29) when both distributions are non-normal. To do this, we find (unique) $\mathbf{q}$ and $\tilde{\mathbf{q}}$ such that $\mathcal{S}_1\mathbf{q} = \mathbf{Q}$ and $\mathcal{S}_1\tilde{\mathbf{q}} = \tilde{\mathbf{Q}}$, using the fact that $x + \sum m^{-k/2} q_k(x)$ $x + \sum m^{-k/2} \tilde{q}_k(x)$ are Cornish-Fisher expansions. Then the expansion $x + \sum m^{-k/2} p_k(x)$ is just the formal composition of the second of the above Cornish-Fisher expansions with the inverse of the first (note that this fails if $q > 1$, because $\mathcal{P}_G^q$ is not closed under composition). This means that we can replace (23) by

$$B_k = \sum_{i+j=2k; i>0, j>0} \int_{\mathbb{R}} (\tilde{q}_i(x) - q_i(x))(\tilde{q}_j(x) - q_j(x)\phi(x)dx$$

and in particular $B_2 = \int_{\mathbb{R}} \{2(\tilde{q}_1(x) - q_1(x))(\tilde{q}_3(x) - q_3(x) + (\tilde{q}_2(x) - q_i(x))^2\}\phi(x)dx$, from which we can recover (29).

Monotonicity questions

Problem 7.20 in [11] asks whether, if X and $\tilde{X}$ are random vectors with zero mean, and $Y_m, \tilde{Y}_m$ are defined as above, then $\mathbb{W}_2(Y_m, \tilde{Y}_m)$ must be decreasing as a function of m. A counterexample, in which X and $\tilde{X}$ are integer-valued random variables with finite range, was given in [10]. Here we use the asymptotic expansion to give a partial positive result under some conditions on the distributions of X and $\tilde{X}$.

Proposition 6. *Suppose X and $\tilde{X}$ are $\mathbb{R}^q$-valued random variables with zero mean and satisfy the CC condition. Suppose also that $\tilde{X}$ has all moments finite, while X satisfies the stronger moment condition $\sup_{k\in\mathbb{N}} k^{-1}(\mathbb{E}|X|^k)^{1/k} < \infty$.*

Then there exists m_0 such that $\mathbb{W}_2(Y_{m+1}, \tilde{Y}_{m+1}) \le \mathbb{W}_2(Y_m, \tilde{Y}_m)$ for all $m > m_0$.

Proof. We denote the variances of X and $\tilde{X}$ by Σ and $\tilde{\Sigma}$, and define $A, \mathbf{S}, \tilde{\mathbf{S}}, \mathbf{p}$ as before. We distinguish two cases, depending on whether $\tilde{X}$ and AX have identical moments or not:

Case 1. $\mathbb{E}(\tilde{X}^\alpha) = \mathbb{E}((AX)^\alpha)$ for all multi-indices α. Then by the moment condition on X, we see that the characteristic function χ_{AX} of AX extends to be analytic in a neighbourhood of $\mathbb{R}^q$ in $\mathbb{C}^q$. Since $\tilde{X}$ has the same moments, the same is true of $\chi_{\tilde{X}}$. Using Taylor expansions about 0, it follows that $\chi_{\tilde{X}} = \chi_{AX}$ in a neighbourhood of 0 in $\mathbb{C}^q$, and hence by analytic continuation $\chi_{\tilde{X}} = \chi_{AX}$ on all of $\mathbb{R}^q$. It follows that $\tilde{X}$ has the same distribution as AX, which in turn implies that, for each m, $\tilde{Y}_m$ has same distribution as AY_m.

Then, noting that Y_m has mean 0 and covariance Σ, we deduce from (14) that $\mathbb{W}_2(Y_m, \tilde{Y}_m) = \mathbb{W}_2(Y_m, AY_m) = \mathrm{tr}((A - I)^2\Sigma)$ which does not depend on m, and the result follows.

Case 2. When case 1 does not apply, there is some multi-index α such that $\mathbb{E}((AX)^\alpha) \neq \mathbb{E}(\tilde{Y}^\alpha)$. This means that for some $k \in \mathbb{N}$, $\tilde{S}_k(Ax)$ is not identically equal to $S_k(x)$, and hence p_k is not identically 0. Taking the smallest k for which this holds, we see from (23) that $B_k > 0$, while $B_j = 0$ for $1 \leq j < k$. Then we have

$$\mathbb{W}_2(Y_m, \tilde{Y}_m)^2 = B_0 + B_k m^{-k} + B_{k+1} m^{-k-1} + O(m^{-k-2})$$

from which we deduce that $\mathbb{W}_2(Y_m, \tilde{Y}_m)^2 - \mathbb{W}_2(Y_{m+1}, \tilde{Y}_{m+1})^2 = kB_k m^{-k-1} + O(m^{-k-2})$ which, since $B_k > 0$, is positive for m large enough, as required. □

We remark that the 'stronger moment condition' in the hypotheses is equivalent to the requirement that the moment generating function of X be finite in a neighbourhood of 0 in $\mathbb{R}^q$.

It is natural to ask to what extent the hypotheses of the above proposition are necessary. By approximating the (integer-valued) random variables in the example from [10], one can get an example where X and $\tilde{X}$ are bounded random variables with smooth densities, but $\mathbb{W}_2(Y_2, \tilde{Y}_2) < \mathbb{W}_2(Y_3, \tilde{Y}_3)$, showing that the requirement $m > m_0$ cannot be omitted. Also, by modifying the construction in [10], one can construct bounded integer-valued random variables X and $\tilde{X}$ such that $\mathbb{W}_2(Y_m, \tilde{Y}_m) \geq m^{-1/2}$ for m odd but $\mathbb{W}_2(Y_m, \tilde{Y}_m) = O(m^{-1})$ for m even, which shows that the CC assumption cannot be omitted. One can ask, however, whether it is enough to assume that one of X and $\tilde{X}$ satisfies CC. Related to this is the question mentioned in [10], whether monotonicity (for all m) holds when $\tilde{X}$ is normal with the same covariance as X. Finally one can ask whether the moment conditions can be relaxed.

References

[1] R. N. Bhattacharya and R. R. Rao, *Normal Approximations and Asymptotic Expansions*, SIAM 2010.

[2] A. M. Davie, Pathwise approximation of stochastic differential equations using coupling, preprint: www.maths.ed.ac.uk/~adavie/coum.pdf

[3] P. Hall, *The Bootstrap and Edgeworth Expansion*, Springer-Verlag 1992.

[4] A. Stuart and K. Ord, *Kendall's Advanced Theory of Statistics*, Volume 1, 6th edition, Edward Arnold 1994.

[5] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag 1999.

[6] V. V. Petrov, *Sums of Independent Random Variables*, Springer-Verlag 1975.

[7] S. T. Rachev and L. Ruschendorff, *Mass Transportation Problems*, Volume 1, Theory; Volume 2, Applications. Springer-Verlag 1998.

[8] E. Rio, Upper bounds for minimal distances in the central limit theorem, *Ann. Inst. Henri Poincaré Probab. Stat.* **45** (2009), 802-817.

[9] E. Rio, Asymptotic constants for minimal distances in the central limit theorem, Elect. Comm. in Probab. **16** (2011), 96-103.

[10] W. Schachermayer, U. Schmock and J. Teichmann, Non-monotone convergence in the quadratic Wasserstein distance, *Springer Lecture Notes in Math.* **1979** (2009), 131-136.

[11] C. Villani, *Topics in Optimal Transportation*, AMS 2003.

[12] X. Zhang, A multi-dimensional central limit bound and its application to the Euler approximation of Lévy SDEs, arXiv:1609.08037v2.