Chapter 4

Stochastic gauges

The non-uniqueness of representations $P(C)$ leads to many potential sets of stochastic equations representing the same single master equation. In this chapter, the stochastic gauge formalism is introduced, which systematically characterizes the freedoms available in the stochastic equations after the physical model has been completely specified. The term "stochastic gauges" is used here because of an analogy with electromagnetic gauges: In both cases the equations contain arbitrary (in principle) functions, which have no effect on physical observables, but can have a great influence on the ease with which the calculation proceeds. In the case of stochastic gauges, the choice of gauge has no effect[1] on observable quantum expectation values (3.14) calculated in the limit of an infinite number of samples. For a finite number of samples, however, different gauges affect how rapidly this large sample limit is approached. It will be shown that an appropriate gauge choice can improve simulation efficiency by many orders of magnitude, and also correct biases caused by pathological distributions.

During the derivation of stochastic Langevin equations (3.45) for the variables C_j from a master equation for the density matrix, there are two distinct places where stochastic gauges can be introduced. Firstly, null differential identities on the kernel $\widehat{\Lambda}$, multiplied by arbitrary functions $\mathcal{F}$ and integrated over $P(C)$ can be introduced into the master equation at the point (3.41), which leads to $\mathcal{F}$ finding its way into the

[1] In principle. In practice there is also the issue of boundary term errors, which may be present or not, depending on the gauge chosen. This is discussed in detail in Chapter 6.

FPE, and into the final stochastic equations. These "kernel stochastic gauges" are described in detail in Sections 4.1 to 4.3. Secondly, the diffusion matrix of an FPE does not uniquely specify the noise behavior of the resulting stochastic equations. These degrees of freedom can also be used to introduce arbitrary functions g into the noise matrices B, as described in detail in Section 4.4. A partly restricted set of kernel and diffusion gauges, which will be here dubbed the "standard" gauges, and are sufficient for most purposes, are summarized in Section 4.5.

As in the previous Chapter 3, the derivation is kept as general as possible with the aim of sifting out those elements that are truly necessary for stochastic gauges to be present and to be useful. Several authors have recently (all starting in 2001) proposed phase-space distribution methods that make use of the freedoms that are here dubbed stochastic gauges. This includes the "noise optimization" of Plimak, Olsen, and Collett[2], the "stochastic wavefunction" method of Carusotto, Castin, and Dalibard[1, 65], and the "stochastic gauges" of Deuar, Drummond, and Kheruntsyan[3, 66, 56, 61]. All these are unified within the formalism presented in this chapter, which is itself a generalization of some ideas to be found in the articles by Deuar and Drummond[3] and [56].

4.1 Generalized kernel stochastic gauges

Whenever there are null differential identities on the kernel of the general form

$$\left\{ J^{(0)}(C) + \sum_j J_j^{(1)}(C)\frac{\partial}{\partial C_j} + \frac{1}{2}\sum_{jk} J_{jk}^{(2)}(C)\frac{\partial^2}{\partial C_j \partial C_k} \right\} \widehat{\Lambda}(C) = \mathcal{J}\left[\widehat{\Lambda}(C)\right] = 0, \quad (4.1)$$

a stochastic gauge can be introduced in the following manner:

Since (4.1) is zero, so is its integral when multiplied by non-divergent functions. In particular,

$$\int P(C,t)\mathcal{F}(C,t)\mathcal{J}\left[\widehat{\Lambda}(C)\right] dC = 0 \quad (4.2)$$

with the function $\mathcal{F}(C,t)$ otherwise unspecified. Now one adds zero (in the form (4.2)) to the right hand side of the master equation (3.41) in the derivation of the Fokker-Planck equation of Section 3.4.1. (Actually, to be fully rigorous, one must make sure that the full integral $\int P(\cdot)\widehat{\Lambda}$ over the entire master equation (3.41) plus all "zero" terms (4.2) is convergent, otherwise boundary term errors may result. This issue is considered in detail in Chapter 6, so for now let us defer this and assume that no boundary term errors are present.)

Adding (4.2) introduces an additional master equation term that can be written in the notation of Section 3.4.1 as

$$\mathcal{T}_{\mathcal{J}}\left[\widehat{\Lambda}\right] = \mathcal{F}\mathcal{J}\left[\widehat{\Lambda}\right]. \tag{4.3}$$

Following now the procedure of Sections 3.4.1 and (3.4.2), the result is that the coefficients in the Langevin equations (3.45) become modified to

$$A_j(C) \to \quad A_j^{(\mathcal{F})}(C) = \quad A_j(C) + \mathcal{F}(C)J_j^{(1)}(C) \tag{4.4a}$$

$$D_{jk}(C) \to \quad D_{jk}^{(\mathcal{F})}(C) = \quad D_{jk}(C) + \mathcal{F}(C)J_{jk}^{(2)}(C). \tag{4.4b}$$

In this manner an arbitrary function $\mathcal{F}$ has been introduced into the stochastic equations. The number of these is unlimited provided we have at least one identity of the form (4.1), although there will be at most as many gauges with distinct effects on the equations as we have distinct identities.

The whole procedure is based on properties (4.1) of the kernel, and so the kernel gauges that can be used in a particular simulation will depend on the local subsystem bases used to expand the density matrix, and on how the kernel is constructed out of these.

Some (though certainly not all) broad classes of gauges that are possible with various kernels include:

- Analytic: The procedure used in Section 3.4.3 to ensure a positive diffusive propagator for representations based on kernels analytic in complex variables z_j can be interpreted as using a gauge of the form

$$\left(\frac{\partial}{\partial \mathrm{Re}\left\{z_j\right\}} + i\frac{\partial}{\partial \mathrm{Im}\left\{z_j\right\}}\right)\widehat{\Lambda}(z_j) = 0. \tag{4.5}$$

  This applies for any kernel analytic in z_j.

- Weighting: If constant terms are present during the derivation of an FPE as in (3.43) in Section 3.4.1, they can be converted to deterministic evolution of global weights, provided the kernel contains a global factor Ω, using an identity of the form

$$\left(\Omega\frac{\partial}{\partial\Omega} - 1\right)\widehat{\Lambda} = 0. \tag{4.6}$$

See Section 4.2, below.

- Drift: One can modify the deterministic part of the stochastic equations (3.45) in principle at will, provided compensation is made in the form of appropriate global trajectory weights. Building on the weighting gauge identity (4.6) for kernels with global weights Ω, the identities

$$\left(\Omega\frac{\partial}{\partial\Omega} - 1\right)\frac{\partial}{\partial C_j}\widehat{\Lambda} = 0. \tag{4.7}$$

follow for any variable C_j, and can be used for this purpose. See Section 4.3.

- Reduction: For some kernels, identities can be found that reduce second order partial derivatives of the kernel to first order. These can be used at the level of expressions (3.40) for each part $\mathcal{T}_l$ of the master equation to reduce partial derivative terms obtained with the operator identities (3.39) to lower order. This will result in a change from diffusion behavior in the FPE to deterministic drift, or even better, in transformation of third- or higher-order partial derivative terms (which prevent one from obtaining any FPE) to lower-order terms, which can form part of an FPE. The identities required are of the general form

$$\left(\frac{\partial}{\partial C_j}\frac{\partial}{\partial C_k} - f(C)\frac{\partial}{\partial C_p}\right)\widehat{\Lambda} = 0. \tag{4.8}$$

for some variables C_j, C_k, and C_p, and some function $f(C)$. An example can be found in the recent work of Drummond and Corney on boson and fermion phase-space distributions[67, 68]

4.2 Weighting stochastic gauges

Consider the situation where during the derivation of the FPE in Section 3.4.1, one encounters nonzero constant terms $\sum_l T_l^{(0)}(C)P(C)$. Most typically, this situation occurs with un-normalized density matrices e.g. in the thermodynamic equation (2.29). As it stands, such an equation of the form (3.43) does not lead immediately to stochastic equations for the variables C. However, if the kernel contains a global weight factor Ω, or by a simple modification is given one, this constant term can be converted to a deterministic evolution of this weight.

What is needed is a kernel of the form

$$\widehat{\Lambda}(C) = \Omega\underline{\widehat{\Lambda}}(\underline{C}). \tag{4.9}$$

where the "base" kernel $\underline{\widehat{\Lambda}}(\underline{C})$, containing all the local basis operators for subsystems does not depend on the global weight Ω. For notational convenience let us define the logarithm of the weight by

$$\Omega = e^{C_0}, \tag{4.10}$$

then the full parameter set is $C = \{C_0, \underline{C}\}$. Directly from (4.9), one can see that the gauge identity

$$\left(\frac{\partial}{\partial C_0} - 1\right)\widehat{\Lambda}(C) = 0. \tag{4.11}$$

applies. If, now, one adds "zero" defined as

$$\sum_l \int P(C) T_l^{(0)}(\underline{C}) \left(\frac{\partial}{\partial C_0} - 1\right)\widehat{\Lambda}(C) dC = 0 \tag{4.12}$$

to the master equation in the form (3.41), the original terms in $T_l^{(0)}$ vanish, to be replaced by first order differential terms

$$\sum_l \int P(C) T_l^{(0)}(\underline{C}) \frac{\partial}{\partial C_0}\widehat{\Lambda}(C) dC. \tag{4.13}$$

This can be incorporated into the formalism of the derivation in Section 3.4.1 by including labels j, $k = 0$ in the sums $\sum_j$, $\sum_k$, and noting that now

$$T_{l0}^{(1)}(\underline{C}) = T_l^{(0)}(\underline{C}) \tag{4.14a}$$

$$T_{l0k}^{(2)} = T_{lj0}^{(2)} = T_{l00}^{(2)} = 0. \tag{4.14b}$$

This then follows through to additional drift and diffusion coefficients (3.44) in the FPE (3.37)

$$A_0(\underline{C}) = \sum_l T_l^{(0)}(\underline{C}) \tag{4.15}$$

$$D_{j0} = D_{0j} = D_{00} = 0. \tag{4.16}$$

The new equation for $C_0 = \log\Omega$ becomes

$$dC_0(t) = A_0(\underline{C}, t)\, dt \tag{4.17}$$

(with no noise), while the stochastic equations for all the remaining variables C_j remain unchanged.

Lastly, the global weight can be made complex with no formal change to the above, if one needs to deal with complex constant terms $T_l^{(0)}$.

4.3 Drift stochastic gauges

4.3.1 Mechanism

Consider again a kernel with global weight $\Omega = e^{C_0}$, defined as (4.9) in the previous section. The weighting gauge identity (4.11) implies further gauge identities involving each of the variables in C (including C_0 itself):

$$\begin{aligned} 0 &= \left(\frac{\partial}{\partial C_0} - 1\right)\frac{\partial}{\partial C_j}\widehat{\Lambda}(C) && (4.18a)\\ &= \left(\frac{1}{2}\frac{\partial}{\partial C_0}\frac{\partial}{\partial C_j} + \frac{1}{2}\frac{\partial}{\partial C_j}\frac{\partial}{\partial C_0} - \frac{\partial}{\partial C_j}\right)\widehat{\Lambda}(C) && (4.18b)\\ &= \mathcal{J}_j\left[\widehat{\Lambda}(C)\right]. && (4.18c)\end{aligned}$$

While (4.11) was used to convert constant terms in the "FPE-like" expression (3.43) to first order derivative terms (and hence deterministic evolution in the weight), the new identities can be used to convert first to second order terms. This allows conversion of deterministic evolution in a variable C_j to stochastic changes in the weight. Such a conversion can be highly desirable as will be shown in Chapters 6 and later.

The procedure is similar to that employed in Section 4.2, however some modifications are made to make the final expression for the stochastic equations in C_j more revealing. If one adds "zero", defined this time as

$$\sum_j \int P(C)\mathcal{F}_j(C)\mathcal{J}_j\left[\widehat{\Lambda}(C)\right] dC = 0, \tag{4.19}$$

with arbitrary gauges $\mathcal{F}_j$ for each variable C_j, then in the formalism of Section 3.4.1 additional terms $\int P(C)\mathcal{T}_0\left[\widehat{\Lambda}(C)\right]dC = 0$ are introduced into the master equation, with the superoperator $\mathcal{T}_0$ containing coefficients

$$\begin{aligned} T^{(1)}_{0j}(C) &= -\mathcal{F}_j(C) && (4.20a)\\ T^{(2)}_{0\underline{j}0}(C) &= T^{(2)}_{00\underline{j}}(C) = \mathcal{F}_{\underline{j}}(C) && (4.20b)\\ T^{(2)}_{000}(C) &= = 2\mathcal{F}_0(C), && (4.20c) \end{aligned}$$

where underlined indices have been used to indicate labels for the "base" variables such that $\underline{j} = 1, 2, \dots,$ etc. Following the procedure through, the modified FPE coefficients

$$\begin{aligned} A_j &= \underline{A}_j - \mathcal{F}_j && (4.21a)\\ D_{\underline{j}\,\underline{k}} &= \underline{D}_{\underline{j}\,\underline{k}} && (4.21b)\\ D_{\underline{j}\,0} &= \underline{D}_{\underline{j}\,0} + \mathcal{F}_{\underline{j}} = \underline{D}_{0\underline{j}} && (4.21c)\\ D_{00} &= 2\mathcal{F}_0 && (4.21d) \end{aligned}$$

are obtained (the original coefficients when $\mathcal{F}_j = 0$ are the underlined). The modifications to the deterministic evolution of variables are seen to be $-\mathcal{F}_j\,dt$.

A difficulty with the forms (4.21) is that in general the form of the diffusion matrix elements B_{jk} that actually go into the stochastic equations will depend in some complicated manner on all the $\mathcal{F}_j$. If the aim is to modify only the deterministic (drift) evolution of the variables $C_{\underline{j}}$ (as will be the case in later chapters), one may end up with the side-effect of additional changes in the noises $B_{\underline{j}k}\,dW_k$. This can prevent a clear assessment of what practical effect the change in the equations due a particular gauge will have.

A convenient form to have the noise matrices B in would be to have no change in the base variables $C_{\underline{j}}$ apart from the modification of the drift, and define new (also arbitrary) gauge functions $\mathcal{G}_{\underline{j}}(\mathcal{F})$ that give the exact stochastic terms for the weight variable C_0, and do not introduce any more independent noises than were had in the original formulation. That is, if the prior weight evolution had no diffusion, one can try the ansatz

$$B_{\underline{j}\,\underline{k}} = \underline{B}_{\,\underline{j}\,\underline{k}} \tag{4.22a}$$

$$B_{0\underline{k}} = \mathcal{G}_{\underline{k}}. \tag{4.22b}$$

Using $D = BB^T$, the diffusion matrix elements that would result from (4.22) are

$$D_{\underline{j}\,\underline{k}} = \sum_{\underline{p}} \underline{B}_{\,\underline{j}\,\underline{p}}\,\underline{B}_{\,\underline{k}\,\underline{p}} = \underline{D}_{\,\underline{j}\,\underline{k}} \tag{4.23a}$$

$$D_{\underline{j}0} = \sum_{\underline{p}} \underline{B}_{\,\underline{j}\,\underline{p}}\mathcal{G}_{\underline{p}} = D_{0\underline{j}} \tag{4.23b}$$

$$D_{00} = \sum_{\underline{p}} \mathcal{G}^2_{\underline{p}}. \tag{4.23c}$$

Comparing these with (4.21), one identifies

$$\mathcal{F}_0 = \frac{1}{2}\sum_{\underline{p}} \mathcal{G}^2_{\underline{p}} \tag{4.24a}$$

$$\mathcal{F}_{\underline{j}} = \sum_{\underline{p}} \underline{B}_{\,\underline{j}\,\underline{p}}\mathcal{G}_{\underline{p}}. \tag{4.24b}$$

Note that there has to be a small restriction of the gauge freedoms from (4.21) to (4.24) (one less arbitrary function due to (4.24a)) to achieve the convenient noise matrix B given by the ansatz (4.22).

The final stochastic Langevin equations with gauges $\mathcal{G}_{\underline{k}}$ then are (noting that $\underline{B}_{\,\underline{j}\,\underline{k}} = B_{\underline{j}\,\underline{k}}$)

$$dC_{\underline{j}} = \underline{A}_{\,\underline{j}}\,dt + \sum_{\underline{k}} \underline{B}_{\,\underline{j}\,\underline{k}}\left(dW_{\underline{k}} - \mathcal{G}_{\underline{k}}\,dt\right), \tag{4.25a}$$

$$dC_0 = \underline{A}_0\,dt + \sum_{\underline{k}} \mathcal{G}_{\underline{k}}\left(dW_{\underline{k}} - \frac{1}{2}\mathcal{G}_{\underline{k}}\,dt\right). \tag{4.25b}$$

The (Ito) equation for the actual weight $\Omega = e^{C_0}$ takes on a simpler form

$$d\Omega = \Omega\left\{\underline{A}_o\,dt + \sum_{\underline{k}} \mathcal{G}_{\underline{k}}dW_{\underline{k}}\right\}. \tag{4.26}$$

with no new drift terms. Recapping, several assumptions beyond the identities (4.18) have been made to obtain (4.25), and these are:

1. The integral $\int P(\cdot)\widehat{\Lambda}dC$ for the master equation in form (3.41), including terms containing the $\mathcal{F}_j$, is convergent.

2. The function $\mathcal{F}_0$ is determined by (4.24a), rather than being an independent arbitrary function.

3. The prior evolution of the weight contains no stochastic terms ($\underline{D}_{\underline{j}0} = \underline{D}_{00} = 0$).

Lastly, as with weighting gauges, if one is dealing with a kernel analytic in constant variables z_j, the above derivation goes through for z_j instead of C_j with no formal change, provided the weight variable $C_0 = z_0 = \log\Omega$ is now also complex. The gauges $\mathcal{G}_{\underline{k}}$ would then be arbitrary *complex* functions.

4.3.2 Real drift gauges, and their conceptual basis

If one arbitrarily adjusts the drift behavior of the stochastic equations, there must be some kind of compensation to ensure continued correspondence to the original quantum master equation. When the gauges $\mathcal{G}_j$ are real, the modification of the weight can be understood in terms of compensation for the increments $\underline{B}_{jk}(dW_k - \mathcal{G}_k)$ no longer being described a Gaussian noise of mean zero (in the many infinitesimal timesteps limit when the CLT applies). Let us first look at the simple case of one subsystem with one complex variable z, with the gauge-less weight is constant, and the drift gauge $\mathcal{G}$ real. The equations for the real and imaginary parts of the variables $z = z' + iz''$ and $z_0 = \log\Omega = z_0' + iz_0''$ are

$$
\begin{aligned}
dz' &= \underline{A}'dt + \underline{B}'dW - \underline{B}'\mathcal{G}dt && (4.27a)\\
dz'' &= \underline{A}''dt + \underline{B}''dW - \underline{B}''\mathcal{G}dt && (4.27b)\\
dz_0' &= -\frac{1}{2}\mathcal{G}^2dt + \mathcal{G}dW && (4.27c)\\
dz_0'' &= 0, && (4.27d)
\end{aligned}
$$

where we have used the shorthand $\underline{B}' = \mathrm{Re}\{\underline{B}\}$ and $\underline{B}'' = \mathrm{Im}\{\underline{B}\}$ etc. for all quantities, and omitted writing subscripts $j, k = 1$.

Consider a very small time step dt in which the variables change from $z(t)$ to $z(t + dt)$ etc. This time step is small enough so that the stochastic equations for infinitesimal dt apply, while large enough that the Wiener increment dW is Gaussian distributed due to the CLT. Suppose also that we consider just one particular trajectory so that the initial distributions of z and z_0 are delta functions around their initial values. The probability distribution of the Gaussian noise is

$$\Pr(dW) = e^{-dW^2/2dt}, \tag{4.28}$$

and if there is no gauge ($\mathcal{G} = 0$), the dependence of $z'(t+dt)$ on the realization of the noise dW can be inverted to give

$$dW = \frac{z'(t+dt) - z'(t) - \underline{A}'dt}{\underline{B}'}. \tag{4.29}$$

Substituting this into (4.28), we obtain an expression for the probability distribution[2] of $z(t+dt)$:

$$P(z(t+dt)) = \exp\left[-\frac{(z'(t+dt) - z'(t) - \underline{A}'dt)^2}{2(\underline{B}')^2 dt}\right]. \tag{4.30}$$

If we now arbitrarily modify the drift with a nonzero gauge, (4.28) still applies, but the inverted relation $dW(z'(t+dt))$ now becomes

$$dW = \frac{z'(t+dt) - z'(t) - \underline{A}'dt + \mathcal{G}\,\underline{B}'dt}{\underline{B}'}, \tag{4.31}$$

leading to a different probability distribution $P_{\mathcal{G}}(v(t+dt))$ than the "correct" gaugeless (4.30). Precisely:

$$P_{\mathcal{G}}(z(t+dt)) = \exp\left[-\frac{(z'(t+dt) - z'(t) - \underline{A}'dt + \mathcal{G}\,\underline{B}'dt)^2}{2(\underline{B}')^2 dt}\right] \tag{4.32}$$

$$= P(z(t+dt)) \exp\left[-\frac{\mathcal{G}}{2\underline{B}'}\left\{\mathcal{G}\,\underline{B}'dt + 2(z'(t+dt) - z'(t) - \underline{A}'dt)\right\}\right]. \tag{4.33}$$

However, we can recover from this by introducing a global compensating weight Ω for the trajectory, which is always applied in all observable averages so that

$$\Omega P_{\mathcal{G}}(z(t+dt)) = P(z(t+dt)). \tag{4.34}$$

[2] Note that we use only expressions involving z' to characterize the probability distribution of the entire complex variable z. This is because z' and z'' are completely correlated with each other, since only the one real noise dW determines them both. The imaginary part, z'', could have been used just as well.

Using (4.31)[3] we can write this compensating weight from (4.33) as

$$\Omega = \exp\left[\mathcal{G}\left(-\frac{1}{2}\mathcal{G}dt + dW\right)\right]. \tag{4.35}$$

Identifying Ω from the definition of the kernel (4.9) one sees that the drift gauge equations for dz_0 (4.27c) and (4.27d) produce exactly the right global weight to compensate for the arbitrary drift corrections introduced by nonzero $\mathcal{G}$.

For the many-variable case with N_z base variables z_j plus the global complex log-weight z_0, the gauged stochastic equations (4.25) can be written in vector form as

$$d\boldsymbol{z} = \underline{\boldsymbol{A}}\, dt + \underline{B}\,\boldsymbol{dW} - \underline{B}\,\boldsymbol{\mathcal{G}}dt \tag{4.36a}$$

$$dz_0 = -\frac{1}{2}\boldsymbol{\mathcal{G}}^T\boldsymbol{\mathcal{G}}dt + \boldsymbol{\mathcal{G}}^T\boldsymbol{dW}, \tag{4.36b}$$

where all bold quantities denote column vectors of N_z elements labeled by $j = 1, \ldots, N_z$ etc. The probability distribution of a noise realization is

$$\Pr(d\boldsymbol{W}) = \exp\left(-\frac{d\boldsymbol{W}^T d\boldsymbol{W}}{2dt}\right), \tag{4.37}$$

since all noise elements dW_j are independent. With no gauge, $d\boldsymbol{W} = (\underline{B}')^{-1}[d\boldsymbol{z}' - \underline{\boldsymbol{A}}'dt]$, leading to the probability distribution

$$P(\boldsymbol{z}(t+dt)) = \exp\left[-\frac{1}{2dt}\left(d\boldsymbol{z}' - \underline{\boldsymbol{A}}'dt\right)^T[(\underline{B}')^{-1}]^T(\underline{B}')^{-1}\left(d\boldsymbol{z}' - \underline{\boldsymbol{A}}'dt\right)\right], \tag{4.38}$$

while with a gauge the noise can be written

$$d\boldsymbol{W} = (\underline{B}')^{-1}\left[d\boldsymbol{z}' - \underline{\boldsymbol{A}}'dt + \underline{B}'\boldsymbol{\mathcal{G}}dt\right]. \tag{4.39}$$

This leads to a new probability distribution

$$\begin{aligned} P_{\mathcal{G}}(\boldsymbol{z}(t+dt)) &= \exp\left[-\frac{\left(d\boldsymbol{z}' - \underline{\boldsymbol{A}}'dt + \underline{B}'\boldsymbol{\mathcal{G}}dt\right)^T[(\underline{B}')^{-1}]^T(\underline{B}')^{-1}\left(d\boldsymbol{z}' - \underline{\boldsymbol{A}}'dt + \underline{B}'\boldsymbol{\mathcal{G}}dt\right)}{2dt}\right] \\ &= P_{\mathcal{G}}(\boldsymbol{z}(t+dt)) \\ &\quad\times \exp\left[-\frac{\boldsymbol{\mathcal{G}}^T(\underline{B}')^{-1}(d\boldsymbol{z}' - \underline{\boldsymbol{A}}'dt) + (d\boldsymbol{z}' - \underline{\boldsymbol{A}}'dt)^T[(\underline{B}')^{-1}]^T\boldsymbol{\mathcal{G}} + \boldsymbol{\mathcal{G}}^T\boldsymbol{\mathcal{G}}}{2}\right] \\ &= P_{\mathcal{G}}(\boldsymbol{z}(t+dt))\exp\left[-\boldsymbol{\mathcal{G}}^T d\boldsymbol{W} + \frac{1}{2}\boldsymbol{\mathcal{G}}^T\boldsymbol{\mathcal{G}}dt\right] \end{aligned} \tag{4.40}$$

[3]Not (4.29), since that applied only for $\mathcal{G} = 0$.

Where the last line was arrived at using (4.39). This requires a compensating weight

$$\Omega = \exp\left[\boldsymbol{\mathcal{G}}^T\left(-\frac{1}{2}\boldsymbol{\mathcal{G}}dt + d\boldsymbol{W}\right)\right], \tag{4.41}$$

which is given by the gauged equations if $\Omega = e^{dz_0}$.

As corollaries of this:

1. If there was no noise in the original equations ($\underline{B} = 0$), or the noise matrix was singular, then there would be no way a global weight could compensate for an arbitrary change in the drift equations — the new drift would just be plain wrong. For the case of very small noise in the $\boldsymbol{z}$ equations, even small changes with respect to the original drift $\underline{\boldsymbol{A}}$ require very large weight compensation. These are the reasons why the gauges $\boldsymbol{\mathcal{G}}$ are multiplied by the coefficients $\underline{B}$ in the $d\boldsymbol{z}$ equations, but not in the weight equation for dz_0. Thus

   If there is no random component to the evolution of a variable z, its drift cannot be modified using a drift gauge. If the noise matrix is singular, *no* variable drift can be modified using drift gauges.

2. While the weight equation (4.26) appears disordered and random due to the presence of noises dW_k, the evolution of Ω is actually strictly tied to the random walk of the other variables z_j, and acts to exactly compensate for any gauge modifications of the original drift.

3. If we consider the ($2N_z$–dimensional) phase space of all base variables in $\underline{C} = \{z_j\}$, then a single real drift gauge $\mathcal{G}_k$ changes the drift only in the phase-space direction specified by the kth column $\underline{\boldsymbol{B}}_k$ of the noise matrix[4], associated with the kth noise dW_k.

4.3.3 Complex gauges

The change in the drift due to complex gauges $\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{G}}' + i\boldsymbol{\mathcal{G}}''$ is always

$$\boldsymbol{A} \quad = \quad \underline{\boldsymbol{A}}dt - \underline{B}\boldsymbol{\mathcal{G}}'dt - i\underline{B}\boldsymbol{\mathcal{G}}''dt. \tag{4.42}$$

[4]So that $\underline{B} = [\underline{\boldsymbol{B}}_1\ \underline{\boldsymbol{B}}_2\ \ldots]$.

Table 4.1: Tally of drift gauge freedoms and comparison to phase space size.

	Number	Kind
Size of base phase-space $\underline{C} = \{z_{\underline{j}}\}$ (degrees of freedom)	$2N_z$	real
Size of full phase-space $C = \{z_0, \underline{C}\}$	$2N_z + 2$	real
Number of standard drift gauges $\mathcal{G}_k$	N_z	complex
Number of noises dW_k in standard drift gauge scheme	N_z	real
Number of broadening drift gauges $\breve{\mathcal{G}}_j$	N_z	complex
Number of noises in broadening drift gauge scheme (4.84)	$3N_z$	real

Imaginary gauges $\mathcal{G}''_k$, in contrast to real ones, lead to changes in the drift only in a particular direction $i\underline{\boldsymbol{B}}_k$ orthogonal to that affected by a real gauge. Note however, that this is a *particular* direction, and not any of the infinitely many orthogonal to $\underline{\boldsymbol{B}}_k$. This kind of drift modification orthogonal to noise direction is not compensated for by changing the weight of the trajectory $|\Omega|$, but by modification of its phase $e^{iz''_0}$, leading to an interference effect between trajectories. It appears harder to grasp intuitively that this fully compensates, but in simulations such gauges are also seen to preserve observable averages (see Part B).

Similarly to purely real gauges, no drift modification can be made if the noise matrix $\underline{B}$ is singular, as seen from (4.42). Conversely, with non-singular $\underline{B}$ one has, in theory, full freedom to modify the deterministic evolution of the complex z_j to any arbitrary functional form. The weight Ω will keep exactly track of these modifications by virtue of using the same noises $d\boldsymbol{W}$. In fact, the weight evolution can be written as a deterministic function of the evolution of the remaining variables:

$$dz_0 = \boldsymbol{\mathcal{G}}^T \left\{ \underline{B}^{-1} \left(d\boldsymbol{z} - \underline{\boldsymbol{A}}\, dt \right) + \frac{1}{2} \mathcal{G}\, dt \right\}. \tag{4.43}$$

A summary of drift gauge freedoms can be found in Table 4.1.

4.3.4 Weight spread

While the global weight e^{z_0} is completely determined by the evolution of the base variables z_j, in observable calculations (3.14) it appears (particularly in the denominator) as effectively a random variable. Other things being equal, it is certainly desirable to make the spread of the weights small during a simulation.

Variances to consider

In general there are several properties of the weights that may be of relevance. $\text{Re}\{\Omega\}$ appears directly in the denominator average of observable estimates (3.14) via $\text{Tr}\left[\widehat{\Lambda}\right] \propto \Omega$. On the other hand, for general observables, both the phase and magnitude of the weight may be important to calculate the numerator average in an obsrvable estimate (3.14) — the phase may be correlated in important ways with the other variables in the average.

For some observables and kernels, the numerator expression of (3.14) depends mostly only on $\text{Re}\{\Omega\}$. This is particularly so at short times when starting from a state well described as a classical (and separable) mixture of the kernels. In this case, both the numerator and denominator of the observable estimator (3.14) depend mostly on $\text{Re}\{\Omega\}$, and the variance of $\text{Re}\{\Omega\}$ is the most relevant to consider. In more general cases, the variance of $|\Omega|$ will be more relevant.

Log-weight spread estimate

A common hindrance when choosing the gauge is the difficulty of accurately assessing the size of the weight spread that will be produced by a given gauge. Here approximate expressions for the variance of z_0 at small times (4.47) will be derived.

Assuming there is no non-gauge weight drift ($\underline{A}_0 = 0$), the evolution in the variance of $z_0' = \text{Re}\{z_0\}$ is given by

$$d\text{var}\left[z_0'\right] \;=\; d\left\langle (z_0')^2\right\rangle_{\text{stoch}} - 2\left\langle z_0'\right\rangle_{\text{stoch}} d\left\langle z_0'\right\rangle_{\text{stoch}}. \tag{4.44}$$

Using properties of the Ito calculus, one can evaluate the time increments of these quantities from the stochastic equation (4.36b) as

$$d\langle z_0'\rangle_{\rm stoch} = \langle \mathrm{Re}\{A_0\}\rangle_{\rm stoch}\,dt = \frac{1}{2}\sum_k \left\langle (\mathcal{G}_k'')^2 - (\mathcal{G}_k')^2 \right\rangle_{\rm stoch} \tag{4.45a}$$

$$d\left\langle (z_0')^2\right\rangle_{\rm stoch} = \langle 2z_0'\mathrm{Re}\{A_0\}\rangle_{\rm stoch}\,dt + \sum_k \left\langle (B_{0k}')^2 dW_k\right\rangle_{\rm stoch} \tag{4.45b}$$

$$= \sum_k \left\langle z_0'(\mathcal{G}_k'')^2 - z_0'(\mathcal{G}_k')^2 + (\mathcal{G}_k')^2\right\rangle_{\rm stoch}, \tag{4.45c}$$

and so in terms of covariances

$$d\mathrm{var}\left[z_0'\right] = \sum_k \left\langle (\mathcal{G}_k')^2 + \mathrm{covar}\left[(\mathcal{G}_k'')^2, z_0'\right] - \mathrm{covar}\left[(\mathcal{G}_k')^2, z_0'\right]\right\rangle_{\rm stoch}. \tag{4.46a}$$

Similarly one finds

$$d\mathrm{var}\left[z_0''\right] = \sum_k \left\langle (\mathcal{G}_k'')^2 - 2\,\mathrm{covar}\left[\mathcal{G}_k'\mathcal{G}_k'', z_0''\right]\right\rangle_{\rm stoch}. \tag{4.46b}$$

A practical approximation to the variance at short times can be gained by assuming lack of correlations between z_0 and $\mathcal{G}$ (i.e. that the covariances are negligible). These approximations then would be

$$\mathrm{var}\left[z_0'(t)\right] \approx V_0'(t) = \mathrm{var}\left[z_0'(0)\right] + \int_0^t \sum_k \left\langle \mathcal{G}_k'(t')^2\right\rangle_{\rm stoch} dt' \tag{4.47a}$$

$$\mathrm{var}\left[z_0''(t)\right] \approx V_0''(t) = \mathrm{var}\left[z_0''(0)\right] + \int_0^t \sum_k \left\langle \mathcal{G}_k''(t')^2\right\rangle_{\rm stoch} dt'. \tag{4.47b}$$

Of course this is never exact because z_0 *does* depend on the gauges due to its evolution, however for small times, equal starting weights ($\mathrm{var}\left[z_0'\right] = \mathrm{var}\left[z_0''\right] = 0$), and *autonomous* gauges (i.e. $\boldsymbol{\mathcal{G}}$ does not depend explicitly on z_0) the expressions (4.47) are good approximations. They can be used to assess the amount of statistical noise that will be introduced by a particular gauge.

Small time regime for log-weights

How small is "small time" in this context? Clearly times when the covariances are much less than the $\sum_k(\mathcal{G}_k)^2$ terms. From (4.46), this will occur for autonomous gauges and equal starting weights if

$$z_0'(t)\sum_k |\mathcal{G}_k|^2 \ll \sum_k (\mathcal{G}_k')^2, \tag{4.48a}$$

and

$$2z_0''(t)\sum_k \mathcal{G}_k'\mathcal{G}_k'' \ll \sum_k (\mathcal{G}_k'')^2. \tag{4.48b}$$

To obtain limits on these, one needs an approximation to z_0. Since $dW_k \approx \mathcal{O}\left(\sqrt{dt}\right)$, then $\boldsymbol{\mathcal{G}}^T\boldsymbol{dW}$ is the dominant term in the evolution of z_0 at short enough times. This implies

$$z_0'(t) \approx \mathcal{O}\left(\sqrt{dt}\sqrt{\sum_k (\mathcal{G}_k')^2}\right) \tag{4.49}$$

$$z_0''(t) \approx \mathcal{O}\left(\sqrt{dt}\sqrt{\sum_k (\mathcal{G}_k'')^2}\right). \tag{4.50}$$

The root of sum of squares is due to independence of noises dW_k. For the approximations to z_0' to apply one needs

$$\begin{aligned}
\sqrt{dt\sum_k (\mathcal{G}_k')^2} &\gg |\mathrm{Re}\{A_0\}\ dt| = \frac{dt}{2}\left|\sum_k \left[(\mathcal{G}_k')^2 - (\mathcal{G}_k'')^2\right]\right| \\
\sum_k (\mathcal{G}_k')^2 &\gg \frac{dt}{4}\left\{\sum_k \left[(\mathcal{G}_k')^2 - (\mathcal{G}_k'')^2\right]\right\}^2 \le \frac{dt}{4}\left\{\sum_k |\mathcal{G}_k|^2\right\}^2 \\
dt &\ll \frac{4\sum_k (\mathcal{G}_k')^2}{\{\sum_k |\mathcal{G}_k|^2\}^2}.
\end{aligned} \tag{4.51}$$

Given this is the case, z_0' is given by (4.49), and so the condition (4.48a) for (4.47a) to hold becomes

$$\begin{aligned}
\sum_k (\mathcal{G}_k')^2 &\gg \sqrt{dt}\sqrt{\sum_k (\mathcal{G}_k')^2}\sum_k |\mathcal{G}_k|^2 \\
dt &\ll \frac{\sum_k (\mathcal{G}_k')^2}{\{\sum_k |\mathcal{G}_k|^2\}^2}.
\end{aligned} \tag{4.52}$$

Since this agrees with the initial assumption of noise term dominance (4.51), it is the short time condition under which (4.47a) is accurate.

For dz_0'' to be given by (4.49), one needs

$$\begin{aligned}
\sqrt{dt\sum_k(\mathcal{G}_k'')^2} &\gg |\mathrm{Im}\,\{A_0\}\,dt| = dt\left|\sum_k \mathcal{G}_k'\mathcal{G}_k''\right| \\
\sum_k(\mathcal{G}_k'')^2 &\gg dt\left\{\sum_k \mathcal{G}_k'\mathcal{G}_k''\right\}^2 \leq dt\sum_j(\mathcal{G}_j')^2\sum_k(\mathcal{G}_k'')^2 \\
dt &\ll \frac{1}{\sum_k(\mathcal{G}_k')^2}.
\end{aligned} \tag{4.53}$$

The condition for (4.47b) is then

$$\begin{aligned}
\sum_k(\mathcal{G}_k'')^2 &\gg 2\sqrt{dt}\sqrt{\sum_k(\mathcal{G}_k'')^2}\sum_k \mathcal{G}_k'\mathcal{G}_k'' \\
\sum_k(\mathcal{G}_k'')^2 &\gg 4\,dt\left\{\sum_k \mathcal{G}_k'\mathcal{G}_k''\right\}^2 \leq 4\,dt\sum_j(\mathcal{G}_j')^2\sum_k(\mathcal{G}_k'')^2 \\
dt &\ll \frac{1}{4\sum_k(\mathcal{G}_k')^2}.
\end{aligned} \tag{4.54}$$

Again, this agrees with the initial assumption of noise term dominance (4.53), and is the short time condition under which (4.47b) is accurate.

Log-weight variance limits

As will be discussed in Section 7.4 and Appendix A, the variance of z_0' should be $\lesssim \mathcal{O}(10)$ for the simulation to give results with any useful precision. Basically when z_0' of a variance $\mathcal{O}(10)$ or more, the high positive z_0' tail of the distribution contains few samples, but much weight, leading to possible bias. Also if z_0'' has a standard deviation of $\mathcal{O}(\pi)$ or more, mutual cancelling of trajectories with opposite phases will dominate the average, concealing any average in noise for reasonable sample sizes $\mathcal{O}(\lesssim 10^5)$.

This imposes a limit on how long a gauged simulation can last, and so it is extremely desirable to keep the magnitude of the gauges as small as possible.

As a corollary to this point, it is desirable to ensure that all drift gauges are zero at any attractors in phase space, to avoid accumulating unnecessary randomness in the weights when no significant evolution is occuring.

Direct weight spread estimate

One can also investigate the evoultion of the variance of the weight Ω directly, and analytic estimates can be obtained in some situations. To better understand the effect of real or imaginary gauges on the weight, let us consider the case where gauges $\mathcal{G}_k$ and weight Ω are decorrelated, there is no base weight drift ($\underline{A}_0 = 0$), and $\Omega = 1$ initially for all trajectories. Let $\Omega = \Omega' + i\Omega''$, then from (4.26)

$$d\Omega' = \sum_k (\Omega'\mathcal{G}'_k - \Omega''\mathcal{G}''_k) \quad dW_k \tag{4.55a}$$

$$d\Omega'' = \sum_k (\Omega'\mathcal{G}''_k + \Omega''\mathcal{G}'_k) \quad dW_k. \tag{4.55b}$$

If we consider the evolution of the second order moments of the weights, then under the uncorrelated Ω and $\boldsymbol{\mathcal{G}}$ assumption, it can be written as a closed system of equations

$$d\left\langle \begin{bmatrix} [\Omega']^2 \\ [\Omega'']^2 \\ \Omega'\Omega'' \end{bmatrix} \right\rangle_{\text{stoch}} = \begin{bmatrix} c_1 & c_2 & -2c_3 \\ c_2 & c_1 & 2c_3 \\ c_3 & -c_3 & c_1 - c_2 \end{bmatrix} \left\langle \begin{bmatrix} [\Omega']^2 \\ [\Omega'']^2 \\ \Omega'\Omega'' \end{bmatrix} \right\rangle_{\text{stoch}} dt, \tag{4.56}$$

where

$$c_1(t) = \sum_k \left\langle [\mathcal{G}'_k(t)]^2 \right\rangle_{\text{stoch}} \tag{4.57a}$$

$$c_2(t) = \sum_k \left\langle [\mathcal{G}''_k(t)]^2 \right\rangle_{\text{stoch}} \tag{4.57b}$$

$$c_3(t) = \sum_k \langle \mathcal{G}'_k(t)\mathcal{G}''_k(t) \rangle_{\text{stoch}} . \tag{4.57c}$$

This system can be solved, and remembering that $\langle \Omega(t) \rangle_{\text{stoch}} = 1$ (from (4.55)), and $\Omega(0) = 1$, one obtains the solutions

$$\begin{aligned}
\text{var}\,[\Omega'] &= \frac{1}{2} e^{\int_0^t c_1(t')dt'} \left[e^{\int_0^t c_2(t')\,dt'} + e^{-\int_0^t c_2(t')\,dt'} \cos\left(2\int_0^t c_3(t')\,dt'\right)\right] - 1 \\
\text{var}\,[\Omega''] &= \frac{1}{2} e^{\int_0^t c_1(t')\,dt'} \left[e^{\int_0^t c_2(t')\,dt'} - e^{-\int_0^t c_2(t')\,dt'} \cos\left(2\int_0^t c_3(t')\,dt'\right)\right] - 1 \\
\langle \Omega'\Omega'' \rangle_{\text{stoch}} &= \frac{1}{2} e^{\int_0^t [c_1(t') - c_2(t')]\,dt'} \sin\left(2\int_0^t c_3(t')\,dt'\right).
\end{aligned} \tag{4.58}$$

If the gauge averages $c_j(t)$ are approximately constant with time, then at short times, when the condition that Ω and $\mathcal{G}_k$ are uncorrelated holds, the variance of the weight is

$$\mathrm{var}\left[\Omega'\right] = t\sum_k \left\langle (\mathcal{G}'_k)^2\right\rangle_{\mathrm{stoch}} + \frac{t^2}{2}\sum_k \left\langle (\mathcal{G}'_k - \mathcal{G}''{}_k)^2\right\rangle_{\mathrm{stoch}} + \mathcal{O}\left(t^3\right). \tag{4.59}$$

In similar manner, one can also obtain

$$\mathrm{var}\left[\,|\Omega|\,\right] \;=\; \exp\left[\int_0^t \left\langle |\mathcal{G}_k(t')|^2\right\rangle_{\mathrm{stoch}} dt'\right] - 1. \tag{4.60}$$

Some more conslusions about drift gauge forms

For most observables it is the variance of Ω' that is most relevant for uncertainties in the finite sample estimates, because it appears both in the numerator and denominator of (3.14). However, for some like the local quadratures $(\widehat{a}_k^\dagger - \widehat{a}_k)/2i$ the modulus of the weight is more relevant in the numerator of the observable expression (3.14). One can see from (4.59) that at short times imaginary gauges lead to smaller real weight spreads because the variance grows only quadratically, most of the gauge noise going into Ω''. This makes imaginary drift gauges more convenient for short time estimates of most observables

4.4 Diffusion stochastic gauges

The Fokker-Planck equation specifies directly only the diffusion matrix $D(C)$, which is then decomposed via

$$D = BB^T \tag{4.61}$$

into noise matrices B, however these are not specified completely. This freedom in the choice of B, leads to a different kind of stochastic gauge than considered in the previous sections, which will be termed "diffusion gauges" here. No weights are required, and only the noise terms are modified. As with kernel stochastic gauges, the diffusion gauges may in some cases be used to choose a set of equations with the most convenient stochastic properties. The non-uniqueness of B has always been known, but has usually been considered to simply relabel the noises without any useful consequences. It has, however, been recently shown by Plimak *et al* [2] that using a different B than the obvious "square root" form $B = B_0 = D^{1/2}$ leads to impressive improvement in the efficiency of positive P simulations of the Kerr oscillator in quantum optics. (This has a similar form of the Hamiltonian to the nonlinear term in (2.17).) This somewhat surprising result leads us to try to quantify the amount of freedom of choice available in the noise matrices In the process, it will also become apparent why some extra properties of the kernel beyond the most general case are needed for diffusion gauges to be useful for simulations.

4.4.1 Noise matrix freedoms with general and complex analytic kernels

At first glance there appears to be a great deal of freedom in the choice of noise matrix B. Consider that the relationship $D = BB^T$ for N_v real variables can be satisfied by a $N_v \times N_W$ noise matrix, subject to the $N_v(N_v+1)/2$ real constraints $D_{jk} = \sum_p B_{jp}B_{kp}$. The number ($N_W$) of independent real Wiener increments is formally unconstrained. It turns out, however, that much (if not all) of this freedom is simply freedom of labeling and splitting up a single Wiener increment into formally separate parts having no new statistical properties. More on this in Section 4.4.3.

Recaling Section 3.4.2, the Ito Langevin equations (3.45), (3.56), (4.25), or (4.36) can also be written (Whether the C_j are real or complex) as

$$dC_j(t) = A_j(C,t)dt + dX_j(C,t), \tag{4.62}$$

where the $\sum_k B_{jk}dW_k$ terms have been coalesced into a single stochastic increment dX_j. Only the means and variances of the Wiener increments dW_k are specified, and so by the properties of Ito stochastic calculus, the only binding relationships for the stochastic terms dX_j are (3.50) — mutual variances specified by the stochastic average of the diffusion matrix, and zero means.

Let us ask the question "do the relationships (3.50) completely specify the statistical properties of the dX_j?".

Consider firstly that because all the increments are taken to be "practically" infinitesimal in any simulation, then over any significant timescale the noise due to the Wiener increments dW_k will be effectively Gaussian, whatever the actual distribution of dW_k used[5]. This is due to the central limit theorem. Also, (3.50a) is guaranteed by the independence of the B_{jk} and dW_k in the Ito calculus, combined with $\langle dW_k(t)\rangle_{\text{stoch}} = 0$. These constraints then imply that to check if there is any freedom in the statistical properties of the dX_j it suffices to compare the number of covariance conditions (3.50b) with the number of covariance relations for the dX_j.

Seemingly this is trivial: for N_v real variables C_j, there are $N_v(N_v+1)/2$ conditions (3.50b), and the same number of possible covariance relationships $\langle dX_j\, dX_k\rangle_{\text{stoch}}$. This means that

> In the general case of a kernel with no additional symmetry properties, all the (formally different) possible choices of the noise matrix B have exactly the same effect on the statistical behavior of the Langevin equations.

This is despite the noise matrix having formally many free parameters. These then are largely freedoms to relabel and split up the Wiener increments without introducing any new statistical behavior, as will be discussed in Section 4.4.

This simple comparison of conditions and relationships also explains why trying various forms of the noise matrices B was for a long time thought not to have any useful effect — under general conditions it doesn't.

It has been found, however, that a positive P simulation can be optimized by particular choices of B[2]. Let us consider the case when the kernel $\widehat{\Lambda}$ can be written as an analytic function of N_z complex variables $C = \{z_j\}$. Now, there are $N_z(N_z+1)/2$ complex conditions (3.50b). As for statistical properties of the dX_j, we have to consider independently both the real and imaginary parts, dX'_j and dX''_j respectively. Their covariance relations are $\left\langle dX'_j\, dX'_k\right\rangle_{\text{stoch}}$, $\left\langle dX''_j\, dX''_k\right\rangle_{\text{stoch}}$, and $\left\langle dX'_j\, dX''_k\right\rangle_{\text{stoch}}$, numbering $N_z(2N_z+1)$ real relations in all. Since there are two real conditions per one complex, this means that there are left over N_z^2 degrees of freedom in the covariance relations between elements of the complex quantities dX_j. These might be used to tailor these N_z^2 free variances or covariances to our needs.

[5]Provided that the variance of dW_k is finite.

This begs the obvious question of "why do the analytic complex variable kernels seem allow more freedom than a general real variable kernel", since they can also be written in terms of real variables – we just split each complex variable into real and imaginary parts.

The key lies in the fact that kernels analytic in complex variables have special symmetry properties that give us freedom to choose the complex derivatives of the kernel as (3.51), or indeed as

$$\frac{\partial\widehat{\Lambda}(C)}{\partial z_j} = \mathcal{F}\frac{\partial\widehat{\Lambda}(C)}{\partial \text{Re}\{z_j\}} - i(1-\mathcal{F})\frac{\partial\widehat{\Lambda}(C)}{\partial \text{Im}\{z_j\}} \tag{4.63}$$

using an analytic kernel stochastic gauge $\mathcal{F}$. This "analytic" symmetry then enters into the Fokker-Planck equation, and finds its way into the Langevin equations as freedom of choice of noise increments dX_j. As with all stochastic gauges, another way of looking at this is that because the kernel has certain symmetry properties, there is a whole family of distributions $P(C)$ of such kernels corresponding to the same quantum density matrix $\widehat{\rho}$. Choosing gauges chooses between these different distributions.

In the following subsections, several specific types of freedoms (or "stochastic gauges") available in the noise matrix B will be considered, and a tally will be made at the end in Table 4.3.

4.4.2 Standard form of diffusion gauges for analytic kernels

Let us consider the analytic kernel case, where it was seen above that there may be useful noise matrix freedoms. Such kernels tend also to be the most convenient generally because a positive propagator, and so a stochastic interpretation, is always guaranteed by the method in Section 3.4.3.

Since $D = D^T$ is square and can be made symmetric[6], its matrix square root is also symmetric $\sqrt{D} = \sqrt{D}^T$, and can be used as a noise matrix B_0

$$B_0 = \sqrt{D} \tag{4.64}$$

$$D = B_0 B_0^T. \tag{4.65}$$

[6]Since $\partial^2/\partial v_1 \partial v_2 = \partial^2/\partial v_2 \partial v_1$ for any variables v_1 and v_2 can be used in the FPE to achieve $D_{jk} = D_{kj}$.

This square root form can be considered as the "obvious" choice of decomposition, unique apart from the N_z signs of the diagonal terms[7]. However, for any complex orthogonal O such that $OO^T = I$, if B_0 is a valid decomposition of D, then so is the more general matrix $B = B_0 O$. So, any matrix in the whole orthogonal family

$$B = B_0\,O \tag{4.66}$$

is a valid decomposition. A general complex orthogonal matrix can be written explicitly using an antisymmetric matrix basis $\sigma^{(jk)}$, $(j \neq k = 1,\dots,N_z)$ having matrix elements

$$\sigma^{(jk)}_{lp} = \delta_{jl}\delta_{kp} - \delta_{jp}\delta_{kl}. \tag{4.67}$$

With these $(N_z - 1)N_z/2$ independent $N_z \times N_z$ matrices $\sigma^{(jk)}$ the general form is

$$O = \exp\left(\sum_{j<k} g_{jk}(C, C^*, t)\,\sigma^{(jk)}\right). \tag{4.68}$$

The g_{jk} are the $N_z(N_z-1)/2$ complex diffusion gauge functions, which can in principle be *completely arbitrary*, including dependence on all variables in C (not necessarily analytic), and the time variable t, without affecting the correspondence between the Langevin stochastic equations, and the FPE.

As an example, in the case of two complex variables, there is one complex gauge function g_{12}, and the resulting transformation is

$$\begin{aligned} O &= \exp\left(g_{12}\sigma^{(12)}\right) \\ &= \cos(g_{12}) + \sigma^{(12)}\sin(g_{12}), \end{aligned} \tag{4.69}$$

where the anti-symmetric matrix $\sigma^{(12)}$ is proportional to a Pauli matrix:

$$\sigma^{(12)} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{4.70}$$

Hence, e.g. if the diffusion matrix is diagonal, B_0 is also, and the transformed (but equivalent) noise matrix becomes:

$$B = \begin{bmatrix} \sqrt{D_{11}}\cos(g_{12}) & \sqrt{D_{11}}\sin(g_{12}) \\ -\sqrt{D_{22}}\sin(g_{12}) & \sqrt{D_{22}}\cos(g_{12}) \end{bmatrix}. \tag{4.71}$$

[7] One for each complex variable.

The square form of B (4.66) with O given by (4.68) will be termed here the "standard" diffusion gauge, to distinguish it from some other forms that will be discussed in later subsections.

4.4.3 Real standard diffusion gauges and noise mixing

Not all the canonical diffusion gauges g_{jk} are useful. One class of useless gauges to avoid are those that seemingly change B, but effectively only swap around the linear combinations of noises dW_k, without affecting dX_j or hence any statistical properties of the equations. Their main effect is usually to complicate the math.

Consider, for example, the two-complex variable case given by (4.69). For a purely real diffusion gauge $g_{12} = g'_{12}$, the effect of the noise terms in the Langevin equations becomes

$$\boldsymbol{dX} = B(g'_{12})\,\boldsymbol{dW} = B_0 \begin{bmatrix} \cos g'_{12} & \sin g'_{12} \\ -\sin g'_{12} & \cos g'_{12} \end{bmatrix} \begin{bmatrix} dW_1 \\ dW_2 \end{bmatrix} = B_0 \boldsymbol{dW}'. \qquad (4.72)$$

This is just a rotation of the noises $\boldsymbol{dW}$, leading to $\boldsymbol{dW}'$ being a linear combination of the noises dW_1 and dW_2. The new noises dW'_j have the same statistical properties as the old dW_j. Thus $\boldsymbol{dX}(g'_{12}) = B(g'_{12})\,\boldsymbol{dW}$ has exactly the same effect in the equations as the un-gauged $\boldsymbol{dX}(g'_{12} = 0)$, regardless of any complicated form of g'_{12} we choose to try.

More generally, with N_z complex variables, there are at least $N_z(N_z - 1)/2$ such useless noise rotations available, one for each pair of variables. Looking at the standard orthogonal matrix form (4.68), one sees that the terms in the exponential proportional to real parts of gauges g_{jk} represent simply all these useless rotations.

In a formalism with only real variables C_j, and no extra symmetries, diffusion gauges arise formally in the same manner as in Section 4.4.2, although all the matrices D, B, B_0, and O must now be real, making g_{jk} also only real. This means that in such a case *all* the resulting real gauge functions in $O(g_{jk})$ are useless noise mixers, and no useful modifications of the stochastic equations can be achieved by choosing them. This explains why the potential of noise matrix choice to improve the stochastic equations went unnoticed for many years.

This subsection can be summarized as

> Only the imaginary parts of the standard diffusion gauges (4.68) can affect the statistical properties of the stochastic equations.

Lastly, it may be worth pointing out that real gauges can achieve the same useless noise rotations in non-square noise matrices as well. For example, for any arbitrary real function f the noise matrix

$$B^{(f)} = \begin{bmatrix} B\cos f & B\sin f \end{bmatrix} \tag{4.73}$$

satisfies $B^{(f)}[B^{(f)}]^T = D$ just as well as the original B, whatever the form of the arbitrary complex function f. The stochastic increments are (for N_z complex variables)

$$\begin{aligned} dX_j &= \sum_{k=1}^{2N_z} B_{jk}^{(f)} dW_k \\ &= \sum_{k=1}^{N_z} B_{jk}(\cos f dW_k + \sin f dW_{N_z+k}) \\ &= \sum_{k=1}^{N_z} B_{jk} dW'_k, \end{aligned} \tag{4.74}$$

with the combined noises dW'_k having the same statistical properties irrespective of the choice of f.

4.4.4 Imaginary standard diffusion gauges and statistics

The imaginary part of standard diffusion gauges $g''_{jk} = \text{Im}\{g_{jk}\}$ *do* affect the statistics. Again consider the two-variable case of diagonal D of expression (4.69), this time with imaginary gauge $g_{12} = ig''_{12}$.

$$\boldsymbol{dX} = B_0 \begin{bmatrix} \cosh g''_{12} & i\sinh g''_{12} \\ -i\sinh g''_{12} & \cosh g''_{12} \end{bmatrix} \begin{bmatrix} dW_1 \\ dW_2 \end{bmatrix}. \tag{4.75}$$

For example, in the simplest case of diagonal positive real D, one has (using the notation $dX_j = dX'_j + i dX''_j$)

$$dX'_1 = \sqrt{D_{11}} \cosh g''_{12} dW_1 \tag{4.76a}$$

$$dX'_2 = \sqrt{D_{22}} \cosh g''_{12} dW_2 \tag{4.76b}$$

$$dX''_1 = \sqrt{D_{11}} \sinh g''_{12} dW_2 \tag{4.76c}$$

$$dX''_2 = -\sqrt{D_{22}} \sinh g''_{12} dW_1. \tag{4.76d}$$

While the covariance requirements (3.50b) are satisfied, the variances of the dX_j components may vary depending on g''_{12}. The nonzero correlations are

$$\left\langle (dX'_1)^2 \right\rangle_{\text{stoch}} = \cosh^2 g''_{12} \left\langle D_{11} \right\rangle_{\text{stoch}} dt \tag{4.77a}$$

$$\left\langle (dX'_2)^2 \right\rangle_{\text{stoch}} = \cosh^2 g''_{12} \left\langle D_{22} \right\rangle_{\text{stoch}} dt \tag{4.77b}$$

$$\left\langle (dX''_1)^2 \right\rangle_{\text{stoch}} = \sinh^2 g''_{12} \left\langle D_{11} \right\rangle_{\text{stoch}} dt \tag{4.77c}$$

$$\left\langle (dX''_2)^2 \right\rangle_{\text{stoch}} = \sinh^2 g''_{12} \left\langle D_{22} \right\rangle_{\text{stoch}} dt \tag{4.77d}$$

$$\left\langle dX'_1 \, dX''_2 \right\rangle_{\text{stoch}} = -\frac{1}{2} \sinh 2g''_{12} \left\langle \sqrt{D_{11} D_{22}} \right\rangle_{\text{stoch}} dt \tag{4.77e}$$

$$\left\langle dX'_2 \, dX''_1 \right\rangle_{\text{stoch}} = \frac{1}{2} \sinh 2g''_{12} \left\langle \sqrt{D_{11} D_{22}} \right\rangle_{\text{stoch}} dt. \tag{4.77f}$$

Such imaginary gauges only apply when the model is expressed in complex variables z_j with an analytic kernel. A glance at the canonical gauge form (4.68), indicates that there is then one such imaginary gauge $g''_{jk} = \text{Im}\{g_{jk}\}$ for each possible pair of variables.

4.4.5 Non-standard diffusion gauges and further freedoms

In Section 4.4.1 it was seen that there are N_z^2 degrees of freedom[8] for the statistics of the dX_j, but later, in Sections 4.4.3 and 4.4.4 it was shown that there are at most $N_z(N_z - 1)/2$ useful (imaginary) standard gauges $g_{jk} = i g''_{jk}$. Conclusion: there are some more freedoms not included in the standard square noise matrix expression (4.66) and (4.68). The simplest example occurs when there is only one complex kernel parameter z_1: there is one degree of freedom in dX_1, but zero standard gauges.

[8]The case of a kernel analytic in complex variables is being considered all the time.

To see how the extra non-square gauges can enter the picture, let us consider the simplest $N_z = 1$ case. Here there will be three covariance relations for dX_1' and dX_1'', but two conditions (3.50) on these. Firstly, note that while the standard gauged noise matrix $B = B_0 O$ is square in the complex variables z_j, in the real variables z_j' and z_j'' it is a $2N_z \times N_z$ — in this case a 2×1. This leaves no room for a third degree of freedom in $dX_1 = B\boldsymbol{dW}$. There is, however, no particular limit on the number of columns of B or the number of noises in $\boldsymbol{dW}$, provided $BB^T = D$, so let us instead consider a 2×2 noise matrix,

$$B = \begin{bmatrix} b_{\mathrm{rr}} & b_{\mathrm{ri}} \\ b_{\mathrm{ir}} & b_{\mathrm{ii}} \end{bmatrix}, \tag{4.78}$$

where subscripts r and i denote elements relating to the real and imaginary parts of the complex variable z_1. Since there are only three degrees of freedom , we can set one of the elements to zero with no restrictions on the statistical properties of dX_1 (say $b_{\mathrm{ri}} = 0$). The remaining elements must satisfy the conditions (from (3.50))

$$\langle \mathrm{Re}\{D\}\rangle_{\mathrm{stoch}}\, dt = \langle (dX_1')^2 - (dX_1'')^2\rangle_{\mathrm{stoch}} = dt\,\langle b_{\mathrm{rr}}^2 - b_{\mathrm{ir}}^2 - b_{\mathrm{ii}}^2\rangle_{\mathrm{stoch}} \tag{4.79a}$$

$$\langle \mathrm{Im}\{D\}\rangle_{\mathrm{stoch}}\, dt = 2\,\langle dX_1' dX_1''\rangle_{\mathrm{stoch}} = dt\,\langle b_{\mathrm{rr}} b_{\mathrm{ir}}\rangle_{\mathrm{stoch}}\,, \tag{4.79b}$$

leaving room for one arbitrary function (i.e. gauge) out of the three nonzero elements b_{rr}, b_{ri}, or b_{ii}. In the standard noise matrix form from (3.54) one had $b_{\mathrm{ii}} = 0$, so all the noise matrix elements were then completely specified by (4.79), with no room for a gauge.

4.4.6 Distribution broadening gauges

Consider the case of no diffusion $D = 0$. When the kernel is analytic in complex variables, there can actually be nonzero noise matrices, leading to nonzero noise despite no diffusion in the FPE. All that is required is that the complex B obeys $BB^T = 0$, or in terms of stochastic increments $\boldsymbol{dX} = B\boldsymbol{dW}$ that

$$\left\langle \boldsymbol{dX} \cdot \boldsymbol{dX}^T \right\rangle_{\mathrm{stoch}} = D = 0 \tag{4.80a}$$

$$\langle \boldsymbol{dX} \rangle_{\mathrm{stoch}} = 0. \tag{4.80b}$$

A viable noise matrix is, for example,

$$\breve{B} = \left[\begin{array}{cc} \breve{g} & i\breve{g} \end{array} \right], \tag{4.81}$$

where $\breve{g}$ is a $N_z \times N_W$ complex matrix, all of whose elements can be any arbitrary functions (gauges) we like without affecting $\breve{B}\breve{B}^T = D = 0$. There is no limit on the width N_W. For nonzero diffusion, the same kind of gauges can be attached to an existing noise matrix ($BB^T = D$) at one's leisure via

$$\widetilde{B} = \left[\begin{array}{cc} B & \breve{B} \end{array} \right]. \tag{4.82}$$

The new noise matrix also obeys $\widetilde{B}\widetilde{B}^T = D$. Nonzero noises corresponding to zero diffusion in the FPE are possible because the correspondence between stochastic equations and the FPE is exact only in the limit of infinite samples. Thus the conditions (4.80) apply only in the limit of infinitely many trajectories, and $B = 0$ is just one special case in which these conditions are satisfied for every trajectory on its own.

The diffusion gauges $\breve{g}$ will be termed "distribution broadening" here because their effect is to make the distribution of the complex variables z_j broader, while preserving their complex means and mutual correlations. Generally such broadening gauges are not of much use, simply making everything more noisy and reducing precision of observable averages, but there are some situations when this is advantageous.

One example occurs when the noise matrix is singular, or there is no native diffusion in the FPE (i.e. $dX_j = 0$) for a variable z_j. In this situation drift gauges are unable to make modifications to the deterministic evolution. If one, however, uses a broadening gauge to force a nonzero noise matrix, then the usual drift gauge formalism can be used and arbitrary modifications to the deterministic evolution made. The idea of creating additional drift gauges in this manner was first proposed (in a $N_z = 2$ case) by Dowling[69]. Let us see how this proceeds in a general case:

For clarity in the resultant equations it is best to define a diagonal broadening gauge matrix $\breve{g}$ with elements

$$\breve{g}_{jk} = \breve{g}_j \delta_{jk}. \tag{4.83}$$

Adding the broadening gauge to a pre-existing noise matrix B as in (4.81) and (4.82), one obtains a $N_z \times 3N_z$ noise matrix and $3N_z$ independent real noises. Each of these noises can now have a drift kernel gauge attached to it in the manner described in Section 4.3. For the purpose of arbitrary drift manipulation it suffices to introduce N_z complex drift gauges $\breve{\mathcal{G}}_j$ on just N_z of these noises — say on the noise matrix elements $\breve{g}_j$. This leads then to the stochastic equations

$$dz_j = A_j\,dt + \sum_k B_{jk}dW_k + \breve{g}_j(d\breve{W}_j + id\breve{\widetilde{W}}_j - \breve{\mathcal{G}}_j\,dt) \tag{4.84a}$$

$$dz_0 = A_0\,dt - \frac{1}{2}\sum_j \breve{\mathcal{G}}_j^2\,dt + \sum_j \breve{\mathcal{G}}_j d\breve{W}_j. \tag{4.84b}$$

The "broadening noises" $d\breve{W}_j$ and $d\breve{\widetilde{W}}_j$ are independent real Wiener increments just like the dW_k.

The arbitrary gauge modifications to the drift of a variable z_j are

$$-\breve{g}_j\breve{\mathcal{G}}_j\,dt. \tag{4.85}$$

A related situation occurs when the native stochastic increment dX_j for a variable z_j is small. In such a situation, making a significant change in the drift of z_j would require large compensation in the log-weight z_0, leading to rapidly increasing statistical uncertainty or even bias from widely varying trajectory weights. However, if one introduces a broadening gauge such that

$$\begin{aligned} \mathrm{var}\left[\mathrm{Re}\{\breve{g}_j\}\,d\breve{W}_j\right] &\gg \mathrm{var}\left[\mathrm{Re}\{dX_j\}\right] \\ \left\langle \mathrm{Re}\{\breve{g}_j\}^2\right\rangle_{\mathrm{stoch}} &\gg \sum_k \left\langle \mathrm{Re}\{B_{jk}\}^2\right\rangle_{\mathrm{stoch}}, \end{aligned} \tag{4.86}$$

then the drift of $\mathrm{Re}\{z_j\}$ can be modified at smaller weighting cost then if one had used standard drift gauges. Analogously for $\mathrm{Im}\{z_j\}$. In fact, there appears to be a tradeoff here between the amount of noise introduced into the log-weight z_0 (proportional to $\breve{\mathcal{G}}_j$), and the amount of noise introduced directly into the variable whose drift is being modified (proportional to $\breve{g}_j$).

4.4.7 Diffusion from different physical processes

A commonly occurring situation is that several physically distinct processes give separate contributions to the diffusion matrix, e.g.

$$D = \sum_l D^{(l)}. \tag{4.87}$$

Calculating the square-root noise matrix B_0 may, in some cases, give a very complicated expression in this situation. If this is a hindrance in calculations, a much more transparent noise matrix decomposition is possible. One decomposes each diffusion contribution $D^{(l)}$ separately into its own noise matrix as

$$D^{(l)} = B^{(l)}(B^{(l)})^T. \tag{4.88}$$

and then combines them as

$$B = \begin{bmatrix} B^{(1)} & B^{(2)} & \cdots \end{bmatrix}. \tag{4.89}$$

This results in separate noise processes for each diffusion contribution $D^{(l)}$, and formally separate drift gauges (one complex drift gauge per real noise). The benefit of doing this is that the resulting stochastic equations have stochastic terms of a relatively simple form. On the other hand, a possible benefit of doing things the hard way with B_0 a direct square root of the full diffusion matrix D is that diffusion contributions from different processes l may partly cancel, leading to a less noisy simulation.

4.5 Summary of standard gauges

The standard gauge choices that will be used in subsequent chapters can be summarized in vector form by the equations

$$d\boldsymbol{z} = \underline{\boldsymbol{A}}\, dt + \underline{B}_0 O(\{g\})\, (\boldsymbol{dW} - \boldsymbol{\mathcal{G}}\, dt)\,, \tag{4.90a}$$

$$dz_0 = \underline{A}_0\, dt + \boldsymbol{\mathcal{G}}^T \left(\boldsymbol{dW} - \frac{1}{2}\boldsymbol{\mathcal{G}}\, dt \right). \tag{4.90b}$$

This uses a kernel of the form (4.9) proportional to a global weight $\Omega = e^{z_0}$, and analytic in $N_z + 1$ complex variables z_j, z_0. A weighting kernel gauge, standard drift kernel gauges, and standard imaginary diffusion gauges have been used ($g_{jk} = ig''_{jk}$ forming the set $\{g\}$ of arbitrary real-valued gauge functions). Underlined quantities are those obtained from the FPE before introduction of drift gauges, while the bold quantities denote column vectors with an element per base configuration variable $z_{j\neq 0}$. And so, $\underline{\boldsymbol{A}}$ and $\underline{A}_0$ are drift coefficients from the un-gauged FPE, $\boldsymbol{dW}$ are real Wiener increments (with each element dW_j statistically independent for each j and each timestep, of zero mean and variance dt), and $\boldsymbol{\mathcal{G}}$ are arbitrary complex drift gauge functions. $\underline{B}_0 = \sqrt{\underline{D}}$ is the square-root noise matrix form, O is the $N_z \times N_z$ orthogonal matrix given by (4.68), but in this case dependent only on imaginary gauges ig''_{jk}. The gauge freedoms in this standard formulation are summarized in Table 4.2.

Table 4.2: **Tally of diffusion and noise matrix freedoms in the standard formulation** (4.90): Kernel analytic in complex phase-space variables (Section 3.4.3), standard imaginary diffusion gauges (Section 4.4.2), standard drift gauges (Section 4.3).

	Number	Kind
Base phase-space variables $\underline{C} = \{z_j\}$	N_z	complex
All variables $C = \{z_0, \underline{C}\}$	$N_z + 1$	complex
Noises (Wiener increments dW_k)	N_z	real
Drift gauges $\mathcal{G}_k$	N_z	complex
Imaginary diffusion gauges $g_{jk} = ig''_{jk}$	$N_z(N_z-1)/2$	imaginary

Table 4.3: **Tally of diffusion gauge and noise matrix freedoms.** The degrees of freedom counted are always real, not complex.

Number of:	Kernel analytic in N_z complex variables	General kernel in N_v real variables
Real variables	$2N_z$	N_v
Covariance relations between dX_j	$N_z(2N_z+1)$	$N_v(N_v+1)/2$
Covariance constraints on dX_j	$N_z(N_z+1)$	$N_v(N_v+1)/2$
Potentially useful statistical freedoms between dX_j	N_z^2	0
maximum Wiener increments dW_k	∞	∞
standard Wiener increments dW_k	N_z	N_v
Elements in standard square root noise matrix $B_0 = \sqrt{D}$	$2N_z^2$	N_v^2
Elements in extended noise matrix with $2N_z$ real columns	$4N_z^2$	N_v^2
Canonical gauges g_{jk} in (4.68)	$N_z(N_z-1)$	$N_v(N_v-1)/2$
— useless real gauges g'_{jk}	$N_z(N_z-1)/2$	$N_v(N_v-1)/2$
— useful imaginary gauges g''_{jk}	$N_z(N_z-1)/2$	0
Potentially useful non-standard gauges	$N_z(N_z+1)/2$	0
Useless potential noise spawning gauges	∞	∞
Potential broadening gauges in $\breve{g}$	∞	not applicable
Broadening gauges $\breve{g}_j$ in drift gauge scheme (4.84).	N_z	not applicable