- It is not necessary to calculate the Jacobian or make any other approximations of $\mathbf{g}[\cdot]$.
- The prediction stage only consists of standard linear algebra operations (matrix square roots, outer products, matrix and vector summation)
- The number of computations (including an efficient matrix square root algorithm) scales with dimensions at the same rate as linearisation.
- Constraints can be readily incorporated by applying the constraint to each of the projected set $\mathcal{Y}_i$.

In Appendix A we analyse of the performance of the new transformation algorithm against linearisation in detail. It is shown that the most natural framework to use to compare the two algorithms is a Taylor Series expansion of $\mathbf{g}[\cdot]$ evaluated about $\bar{\mathbf{x}}$. Linearisation introduces errors in the mean calculation at second order and in the covariance at the fourth order. The new method, however, yields errors in the mean and covariance which are both of fourth order. Further, on an absolute term-by-term basis, the errors at all higher orders can be made smaller than those introduced by linearisation. In many applications we expect the effects of the lower order terms to be significant, and so the reduction in errors can lead to significant improvements in estimation accuracy. This is demonstrated in a number of examples given below. When the function is discontinuous linearised estimates are almost incapable of capturing this information. If the discontinuity does not lie at the point of linearisation then the estimate does not contain the information. If the linearisation point lies on the discontinuity then the Jacobian matrix may not exist and hence the covariance cannot be predicted. The new filter uses a distribution of points and captures the effects of the discontinuity if it influences a significant proportion of the distribution.

The analysis also reveals the role which is played by κ: it affects the scaling of the fourth and higher moments of the distribution. Thus κ is a convenient parameter for exploiting knowledge (if available) about the higher moments of the given distribution [2]. It is shown that choosing $\kappa = 2$ for a scalar system leads to errors in the mean and covariance which are sixth order. For multi-dimensional systems choosing $\kappa = 3 - n$ minimises the mean squared error up to the fourth order. There is no restriction on the sign of κ but, if κ is negative, then we cannot interpret the distribution of the sigma points as a probability distribution. Further, when κ is negative there is the possibility, as with all approximation algorithms, that the predicted covariance will be non-positive semi-definite. In this case, we can use a modified form of the algorithm for calculating the covariance:

$$\mathbf{P}_{yy} = \frac{1}{2(n+\kappa)} \sum_{i=1}^{2n} [\mathcal{Y}_i - \bar{\mathbf{y}}][\mathcal{Y}_i - \bar{\mathbf{y}}]^T. \tag{26}$$

Both the original and modified algorithms have the property that, as $n + \kappa$ tends to zero, the mean tends to the value obtained by a truncated second order prediction algorithm. Further, the modified algorithm has the useful property that, in this limit, the covariance tends towards that calculated by linearisation[3].

[2] This leads us to the conclusion that the method is applicable for any prior, symmetric, unimodal distribution and not just Gaussian.

[3] This can be contrasted with an alternative approach of the initial intuition which was explored in [13]. Under that scheme no copies of the previously estimated mean are included in the sample set and the sigma points are scaled using a parameter α. In the limit as α tends to infinity this algorithm predicts the same mean and covariance as the EKF. However when $\alpha = 1$, this method estimates the same mean and covariance as that of the new filter but with $\kappa = 0$.