Application of Kähler manifold to signal processing and Bayesian inference

Jaehyung Choi and Andrew P. Mullhaupt

Department of Applied Mathematics and Statistics, SUNY, Stony Brook, NY 11794

Abstract. We review the information geometry of linear systems and Bayesian inference, and the simplification available in the Kähler manifold case. We find conditions for information geometry of linear systems to be Kähler, and the relation of the Kähler potential to information geometric quantities such as $\alpha$-divergence, information distance and the dual $\alpha$-connection structure. The Kähler structure simplifies the calculation of the metric tensor, connection, Ricci tensor and scalar curvature, and the $\alpha$-generalization of the geometric objects. The Laplace-Beltrami operator is also simplified in the Kähler case. One of the goals in information geometry is the construction of Bayesian priors outperforming the Jeffreys prior, which we use to demonstrate the utility of the Kähler structure.

Keywords: Information geometry, Kähler manifold, signal processing, Bayesian inference, Komaki prior, ARFIMA

INTRODUCTION

Kähler manifolds are important in differential geometry, with applications in several different fields such as superstring theory in theoretical physics, and in our interest, information geometry. Barbaresco introduced Kähler manifolds in information geometry [2] and exploited generalized complex manifolds for information geometry [3, 4]. Symplectic and Kähler structures of the divergence function are also revealed [12]. Recently, Choi and Mullhaupt [5] proved the mathematical correspondence between Kähler manifolds and the information geometry of linear systems. Moreover, the implication of the Kähler manifold for Bayesian inference for linear systems is also provided [6].

Kählerian information geometry has several advantages in describing the information geometry of linear systems [5]. First of all, tensor calculation is simpler. Additionally, the $\alpha$-generalization on the tensors is more straightforward because the Riemann tensor is $\alpha$-linear on the complex manifold. Moreover, searching for the superharmonic priors as suggested by Komaki [7] is more efficient because the Laplace-Beltrami operator in the Kähler geometry is much simpler. This simplicity led to the systematic and generic algorithm for the geometric shrinkage priors [6].

In this paper, we provide a review on the recent developments in the applications of the Kähler manifold to information geometry, in particular, the implications to signal processing and Bayesian inference. First, we provide a brief review on the Kähler manifold, then the Kählerian description of the linear systems is introduced. Then we present the application to Bayesian inference.
REVIEW ON KÄHLER MANIFOLD

We review the fundamentals of the Kähler manifold in this section. Let us start with construction of a complex manifold. To extend a real manifold to a complex manifold, the concept of complexification is necessary. Any functions and vectors can be complexified. The complexified coordinate system $\xi = (\xi^1, \cdots, \xi^n) \in \mathbb{C}^n$ of a complex manifold $M$ is given by

$$\xi = \theta + i\zeta$$

where $\theta$ and $\zeta$ are the real coordinate systems of a real manifold $N$ of dimension $n$. The complex manifold $M$ is the complexification of the manifold $N$ denoted by $N^\mathbb{C}$ and it can also be considered as the product manifold $N \times N$. From now on, we work on the complex manifold $M$ of $\dim \mathbb{C}M = n$.

On the tangent plane at point $p$, also denoted by $T_pM$, the basis vectors are given by the real coordinates as

$$\left\{ \frac{\partial}{\partial \theta^1}, \cdots, \frac{\partial}{\partial \theta^n}; \frac{\partial}{\partial \xi^1}, \cdots, \frac{\partial}{\partial \xi^n} \right\}$$

and the cotangent plane $T^*_pM$, which is dual to the tangent plane, is spanned by

$$\left\{ d\theta^1, \cdots, d\theta^n; dz^1, \cdots, dz^n \right\}$$

where $d\theta^i$ and $dz^i$ are the one-forms of the manifold. Since the vectors on the tangent space and the one-forms on the cotangent space are dual to each other, the basis vectors and the one-forms should satisfy the following identities:

$$\langle d\theta^i, \frac{\partial}{\partial \theta^j} \rangle = \delta^i_j, \langle d\theta^i, \frac{\partial}{\partial \xi^j} \rangle = 0$$

$$\langle dz^i, \frac{\partial}{\partial \theta^j} \rangle = 0, \langle dz^i, \frac{\partial}{\partial \xi^j} \rangle = \delta^i_j$$

where $\langle \cdot, \cdot \rangle$ is the inner product and $\delta^i_j$ is the Kronecker delta.

It is also possible to describe the manifold with the complexified coordinate system. First of all, let us introduce the following vectors:

$$\frac{\partial}{\partial \xi^i} = \frac{1}{2} \left( \frac{\partial}{\partial \theta^i} - i \frac{\partial}{\partial \xi^i} \right), \frac{\partial}{\partial \bar{\xi}^i} = \frac{1}{2} \left( \frac{\partial}{\partial \theta^i} + i \frac{\partial}{\partial \xi^i} \right).$$

The tangent space $T_pM$ is spanned by the basis vectors defined above:

$$\left\{ \frac{\partial}{\partial \xi^1}, \cdots, \frac{\partial}{\partial \xi^n}; \frac{\partial}{\partial \bar{\xi}^1}, \cdots, \frac{\partial}{\partial \bar{\xi}^n} \right\}$$

and its dual cotangent space $T^*_pM$ is spanned by

$$\left\{ d\xi^1, \cdots, d\xi^n, d\bar{\xi}^1, \cdots, d\bar{\xi}^n \right\}$$
such that the vectors and the one-forms in the complexified coordinate system satisfy the similar identities:

\[\langle d\xi^i, \frac{\partial}{\partial \xi^j} \rangle = \delta^i_j, \quad \langle d\xi^i, \frac{\partial}{\partial \bar{\xi}^j} \rangle = 0, \quad \langle d\bar{\xi}^i, \frac{\partial}{\partial \xi^j} \rangle = 0, \quad \langle d\bar{\xi}^i, \frac{\partial}{\partial \bar{\xi}^j} \rangle = \delta^i_j.\]

The manifold has an almost complex structure: it is the linear mapping \( J_p : T_p M \rightarrow T_p M \) such that

\[ J_p \frac{\partial}{\partial \xi^i} = i \frac{\partial}{\partial \xi^i}, \quad J_p \frac{\partial}{\partial \bar{\xi}^i} = -i \frac{\partial}{\partial \bar{\xi}^i} \]

and its matrix representation is the following:

\[
J_p = \begin{pmatrix}
i & 0 \\
0 & -i
\end{pmatrix}
\]

where \( I_n \) the identity matrix of dimension \( n \).

A Hermitian manifold is defined as a complex manifold equipped with the metric tensor of the following property:

\[ g_p(J_pX, J_pY) = g_p(X, Y) \]

where \( X, Y \in T_p M \). First of all, the definition of the Hermitian manifold can be represented in terms of the metric tensor as follows:

\[ g_{ij} = g_{i\bar{j}} = 0 \] (1)

where the metric tensor elements with the mixed indices may not vanish. It is always possible to construct a Hermitian manifold from a complex manifold.

The Kähler form is defined as

\[ \Omega_p = g_p(J_pX, Y) \]

where \( X, Y \in T_p M \). It is antisymmetric under the exchange of \( X \) and \( Y \): \( \Omega_p(X, Y) = -\Omega_p(Y, X) \). It is expressed in terms of the metric tensor:

\[ \Omega = ig_{i\bar{j}}dz^i \wedge d\bar{z}^j \]

where \( \wedge \) is the wedge product.

Now, we can define the Kähler manifold. A Kähler manifold is defined as a Hermitian manifold with a closed Kähler form. The closed Kähler two-form, \( d\Omega = 0 \), is written in the metric tensor components

\[ \partial_i g_{jk} = \partial_j g_{ik}, \quad \partial_i g_{kj} = \partial_j g_{ki}. \] (2)

In the metric tensor expression, the geometry is Kähler if and only if the metric tensor satisfies eq. (1) and eq. (2).
One of the most well-known properties in the Kähler geometry is that the metric tensor on the Kähler manifold is given by

\[ g_{ij} = \partial_i \partial_j \mathcal{K} \]  

where \( \mathcal{K} \) is the Kähler potential. All the information on the metric tensor is encoded in the Kähler potential. The nontrivial elements of the Levi-Civita connection also can be expressed with the Kähler potential:

\[ \Gamma_{ij,\bar{k}} = \partial_i \partial_j \partial_{\bar{k}} \mathcal{K} = (\Gamma_{ij,\bar{k}})^* \]

and the other elements of the Levi-Civita connection vanish. The connection with this property is called the Hermitian connection.

Another well-known fact in the Kähler geometry is that the Ricci tensor is calculated from

\[ R_{ij} = -\partial_i \partial_j \log \mathcal{G} \]

where \( \mathcal{G} \) is the determinant of the metric tensor. The lengthy calculation for Riemann curvature tensor can be skipped in the procedure of obtaining the Ricci tensor.

Finally, it is noteworthy that the Laplace-Beltrami operator is represented with

\[ \Delta = 2g^{ij}\partial_i \partial_j \]

and it is much simpler than the Laplace-Beltrami operator for a non-Kähler manifold.

KÄHLER GEOMETRY OF SIGNAL PROCESSING

In this section, we cover Kählerian information geometry for signal processing proposed by Choi and Mullhaupt [5]. A signal filter transforms an input signal \( x \) to an output \( y \) under the following linear relation:

\[ y(w) = h(w; \xi)x(w; \xi) \]

where \( h(w; \xi) \) is a transfer function in frequency domain \( w \). The \( z \)-transformed transfer function of the causal filter is expressed by

\[ h(z; \xi) = \sum_{r=0}^{\infty} h_r(\xi)z^{-r} \]

where \( h_r \) is the \( r \)-th impulse response function of the linear system. We assume that \( h(z; \xi) \) is holomorphic both in \( \xi \) and \( z \).

It is well-known by Amari and Nagaoka [1] that the metric tensor is determined for stationary processes by the spectral density function \( S(z; \xi) = |h(z; \xi)|^2 \). It is also possible to write down the metric tensor in terms of the transfer function on the complexified manifold. The metric tensor can be represented with \( \eta_r \), the coefficient of \( z^{-r} \) in the
log-transfer function,
\[ g_{ij} = \partial_i \eta_0 \partial_j \eta_0 \]  
\[ g_{ij} = \sum_{r=0}^{\infty} \partial_i \eta_r \partial_j \bar{\eta}_r \]  
(6)

(7)

where \( g_{ij} \) and \( g_{ij} \) are the complex conjugates of \( g_{ij} \) and \( g_{ij} \), respectively. It is straightforward that \( \eta_0 = \log h_0 \).

Choi and Mullhaupt [5] provided the condition on the transfer function of linear systems where the information geometry is the Kähler manifold. In the case of causal filter the characterization is as follows.

**Theorem 1.** Given a holomorphic transfer function, the information geometry of a signal filter is the Kähler manifold if and only if \( h_0 \) is a constant in \( \xi \).

**Proof.** If \( h_0 \) is a constant, the metric tensor expressions, eq.(6) and eq. (7), are given by
\[ g_{ij} = g_{ij} = 0, \quad g_{ij} = \sum_{r=1}^{\infty} \partial_i \eta_r \partial_j \bar{\eta}_r \]
i.e. the manifold is Hermitian. Additionally, it is easy to check that the Kähler form is closed.

If the geometry is Kähler, the manifold is Hermitian where \( g_{ij} = g_{ij} = 0 \) for all \( i \) and \( j \). From this Hermitian condition, it is obvious that \( h_0 \) is constant in \( \xi \). \( \square \)

For the linear system, the Kähler potential is the square of Hardy norm of the log-transfer function on the unit disk \( \mathbb{D} \) [5]:
\[ \mathcal{K} = \frac{1}{2\pi i} \oint_{|z|=1} |\log h(z; \xi)|^2 \frac{dz}{z} = ||\log h(z; \xi)||^2_{H^2(\mathbb{D})} \]  
(8)

and the Kähler potential is also related to the 0-divergence. It is identical to the 0-divergence if \( h_0 = 1 \) for the unilateral transfer function.

According to the literature [5], the benefits of the Kählerian information geometry are the followings. First of all, the calculation of geometric tensors and connection is simplified by the Kähler structure. Additionally, the \( \alpha \)-generalization of the tensors are still \( \alpha \)-linear. Finally, it is easier to find the superharmonic priors on the manifold because the Laplace–Beltrami operator is in the simpler form.

We give an example: One of the most interesting linear systems is the fractionally integrated autoregressive moving average (ARFIMA) model. For the ARFIMA\((p, d, q)\) model of \( \xi = (d, \lambda_1, \ldots, \lambda_p, \mu_1, \ldots, \mu_q) \), the transfer function rescaled by the gain is given by
\[ h(z; \xi) = \frac{(1 - \mu_1 z^{-1})(1 - \mu_2 z^{-1}) \cdots (1 - \mu_q z^{-1})}{(1 - \lambda_1 z^{-1})(1 - \lambda_2 z^{-1}) \cdots (1 - \lambda_p z^{-1}) (1 - z^{-1})^d} \]

where \( \lambda_i \) is a pole from the AR part, \( \mu_i \) is a root from the MA part, and \( d \) is a differencing parameter. The poles and the roots are expected to be on the unit disk. By Theorem 1,
it is clear that the information geometry of the ARFIMA model is Kähler. The Kähler potential of the ARFIMA model, also found in the literature [6], is calculated from eq. (8) as

$$\mathcal{K} = \sum_{n=1}^{\infty} \left| d + (\mu_1^n + \cdots + \mu_q^n) - (\lambda_1^n + \cdots + \lambda_p^n) \right|^2$$

and it is bounded above by \((d + p + q)^2 \pi^2/6\). The metric tensor, derived from eq. (3), is represented by

$$g_{ij} = \begin{pmatrix}
\frac{\pi^2}{6} & \frac{1}{\lambda_j} \log (1 - \lambda_j) & -\frac{1}{\mu_j} \log (1 - \mu_j) \\
\frac{1}{\lambda_i} \log (1 - \lambda_i) & \frac{1}{1 - \lambda_i \lambda_j} & 0 \\
-\frac{1}{\mu_i} \log (1 - \mu_i) & 0 & \frac{1}{1 - \mu_i \mu_j}
\end{pmatrix}$$

where the first column and the first row are for the direction of the fractional differencing parameter \(d\). It is easy to find the metric tensor for the pure ARMA model as a submanifold.

The connection is also found from eq. (4) and it is noteworthy that the connections along the differencing parameter direction are all vanishing. The Ricci tensor is also calculated by eq. (5) and it is also vanishing along the \(d\)-direction. The non-vanishing Ricci tensors are only from the pure ARMA directions with the correction term from the mixing between the pure ARMA piece and the fractionally integrated part:

$$R_{ij} = R_{ij}^{ARMA} + R_{ij}^{ARMA-FI}$$

where \(i\) and \(j\) are not 0.

**GEOMETRIC SHRINKAGE PRIORS OF KÄHLERIAN FILTERS**

First, we review the superharmonic priors proposed by Komaki [7]. The difference in risk function between two Bayesian predictive densities from the Jeffreys prior \(\pi_J\) and the superharmonic prior \(\pi_I\) with respect to the density \(p(y|x)\) is given by

$$\mathbb{E}(D_{KL}(p(y|x)||p_{\pi_I}(y|x^{(N)}))|\xi) - \mathbb{E}(D_{KL}(p(y|x)||p_{\pi_J}(y|x^{(N)}))|\xi))$$

$$= \frac{1}{2N^2} g^{ij} \partial_i \log \left( \frac{\pi_I}{\pi_J} \right) \partial_j \log \left( \frac{\pi_I}{\pi_J} \right) - \frac{1}{N^2} \frac{\pi_I}{\pi_J} \Delta \left( \frac{\pi_I}{\pi_J} \right) + o(N^{-2})$$

where \(N\) is the size of samples. If a positive prior function \(\psi = \pi_I/\pi_J\) is superharmonic, the risk function of the Bayesian predictive density \(p_{\pi_I}\) is decreased with respect to \(p_{\pi_J}\), the predictive density from the Jeffreys prior. Comparing with \(p_{\pi_I}\), \(p_{\pi_I}\) is close to \(p(y|x)\) in the Kullback-Leibler divergence. Superharmonic priors for several probability densities and linear systems have been found [7, 10, 11, 5, 6].

A difficulty in Komaki’s idea is that it is non-trivial to test the superharmonicity of the prior function \(\psi\) in general time series models and linear systems with high dimensionality. Although a superharmonic prior for the AR model in an arbitrary dimension was
found by Tanaka [11], no superharmonic priors for the ARMA models and the ARFIMA models were found. Moreover, any systematic algorithm for finding the Komaki priors was not known.

Recently, a generic algorithm for the shrinkage priors of linear systems is introduced when the information geometry of the model is Kähler [6]. In addition to that, the superharmonic priors for more general time series models and signal filters are efficiently found by the algorithm. The following theorem is useful to find the superharmonic prior functions.

**Theorem 2.** On a Kähler manifold, a positive function $\psi = \Psi(u^* - \kappa(\xi, \bar{\xi}))$ is a superharmonic prior function if $\kappa(\xi, \bar{\xi})$ is subharmonic, bounded above by $u^*$, and $\Psi$ is concave decreasing: $\Psi'(\tau) > 0$, $\Psi''(\tau) \leq 0$.

**Proof.** The details of the proof is given in the literature [6].

If we find a positive subharmonic function, we apply Theorem 2 to obtain a superharmonic function and we exploit the superharmonic function as a shrinkage prior function for the prediction as Komaki [7] suggested.

Fortunately, several choices for $\Psi$ and $\kappa$ are already known [6]. The candidates for $\Psi$ are the following:

\[
\Psi_1(\tau) = \tau^a
\]
\[
\Psi_2(\tau) = \log(1 + \tau^a)
\]

where $\tau$ is positive and $0 < a \leq 1$. Moreover, the ansätze for $\kappa$ are found as

\[
\kappa_1 = \mathcal{H}
\]
\[
\kappa_2 = \sum_{r=0}^{\infty} a_r |h_r(\xi)|^2
\]
\[
\kappa_3 = \sum_{i=1}^{n} b_i |\xi^i|^2
\]

where $a_r$ and $b_i$ are positive real numbers. In particular, $\kappa_1$ is the Kähler potential which is intrinsic on the Kähler manifold. By combining $\kappa$ and $\Psi$, it is easy to construct geometric shrinkage priors like

\[
\psi_1 = (u^* - \mathcal{H})^a
\]
\[
\psi_2 = \log(1 + (u^* - \mathcal{H})^a)
\]

which will outperform the Jeffreys prior in the viewpoint of information theory.

**CONCLUSION**

We have reviewed information geometric applications of Kähler manifolds to linear systems and Bayesian inference, and exposed one of the advantages, the simpler Laplace–Beltrami operator, is applicable to Bayesian inference: finding superharmonic priors on the Kähler manifold is quite straightforward, as we have shown for linear systems, in particular the ARFIMA model.
ACKNOWLEDGMENTS

We are grateful to Robert J. Frey and Mike Tiano for useful discussions. We thank Frederic Barbaresco for notifying his notable works on the Kähler geometry and information geometry. We are also thankful to the organizers of MaxEnt 2014.

REFERENCES

1. Amari, S. and Nagaoka, H., Methods of information geometry, Oxford University Press (2000)
2. Barbaresco, F., Information intrinsic geometric flows, AIP Conf. Proc. 872 (2006) 211-218
3. Barbaresco, F., Information Geometry of Covariance Matrix: Cartan-Siegel Homogeneous Bounded Domains, Mostow/Berger Fibration and Fréchet Median, Matrix Information Geometry, Bhatia, R., Nielsen, F., Eds., Springer (2012) 199-256
4. Barbaresco, F., Koszul Information Geometry and Souriau Geometric Temperature/Capacity of Lie Group Thermodynamics, Entropy (2014) 16 4521-4565
5. Choi, J. and Mullhaupt, A. P., Kählerian information geometry for signal processing, arXiv:1404.2006
6. Choi, J. and Mullhaupt, A. P., Geometric shrinkage priors for Kählerian signal filters, arXiv:1408.6800
7. Komaki, F., Shrinkage priors for Bayesian prediction, Ann. Statistics 34 (2006) 808-819
8. Ravishanker, N., Melnick, E. L., and Tsai, C., Differential geometry of ARMA models, Journal of Time Series Analysis 11 (1990) 259-274
9. Ravishanker, N., Differential geometry of ARFIMA processes, Communications in Statistics - Theory and Methods 30 (2001) 1889-1902
10. Tanaka, F. and Komaki, F., A superharmonic prior for the autoregressive process of the second order, Journal of Time Series Analysis 29 (2008) 444-452
11. Tanaka, F., Superharmonic priors for autoregressive models, Mathematical Engineering Technical Reports, University of Tokyo (2009)
12. Zhang, J. and Li, F., Symplectic and Kähler Structures on Statistical Manifolds Induced from Divergence Functions, Geometric Science of Information 8085 (2013) 595-603