A Mathematical Framework for the Kinetic Stress Index (KSI) and its Attribution Analysis

Documentation for the KSI Phase Transition Analyzer (v14.0)

July 21, 2025

Abstract

This document provides a detailed mathematical exposition of the Kinetic Stress Index (KSI) model and its associated attribution analysis. The framework is designed to detect and diagnose systemic market stress by modeling the correlation structure of a specialized asset quartet as a dynamical system. We first detail the construction of the 20-dimensional state vector, which captures both pairwise kinetics and global systemic coherence. We then derive the KSI as a Mahalanobis distance, a geometry-aware measure of statistical anomaly. Finally, we present the methodology for decomposing a stress signal into its constituent sources, enabling a nuanced interpretation of market dislocations. This document serves as a formal companion to the Python implementation.

Contents

1 Core Model Philosophy and Parameters

The KSI framework quantifies market stress by measuring how "surprising" the current state of the market's correlation structure is, relative to its own recent history. It achieves this by combining two complementary perspectives:

- **Pairwise Kinetics (The "Trees"):** Analyzing the individual relationships between asset pairs. We track not only their correlation level (position) but also their rate of change (velocity) and the rate of that change (acceleration). This provides a granular view of the forces acting within the system.
- **Global Dynamics (The "Forest"):** Analyzing the system as a whole. We track a single, powerful measure of market-wide "herding" or coherence—the maximum eigenvalue of the correlation matrix—and its rate of change.

By synthesizing these views into a single state vector, we create a rich snapshot of the market's structural health. The model's key parameters are defined below.

Table 1: Key Symbols and Parameters in the KSI Framework.

Symbol	Definition	Value	Justification for Inclusion
n	Number of assets in the quartet	4	A specialized set chosen to represent the tension between short-term policy and long-term structural fears.
w	Correlation window	60 days	Balances stability against responsiveness to new information in the correlation calculation.
l	Mahalanobis lookback	252 days	Defines the one-year historical baseline of "normal" system behaviour against which anomalies are measured.

2 State Vector Construction ($\mathbf{s}_t \in \mathbb{R}^{20}$)

The 20-dimensional state vector is the mathematical core of the model. Its construction is a multi-step process that transforms raw asset prices into a robust representation of market dynamics.

2.1 Pairwise Kinetics (18 dimensions)

For the $k = \binom{n}{2} = \binom{4}{2} = 6$ unique asset pairs (i, j), we compute their kinetic properties.

1. **Log-returns:** The analysis begins with daily log-returns, which are stationary and represent continuously compounded returns.

$$r_{i,t} = \ln(P_{i,t}/P_{i,t-1})$$

2. **Spearman Correlation ($\rho_{ij,t}$):** The rank correlation is computed over the window w. This non-parametric method is chosen for its resilience to outliers and non-linear relationships common in financial data.

$$\rho_{ij,t} = \text{rank-corr}(r_{i,t-w+1:t}, r_{j,t-w+1:t})$$

3. **Position Vector ($\boldsymbol{\theta}_t \in \mathbb{R}^6$):** To stabilize variance and map the bounded correlations ($[-1, 1]$) to the unbounded real line, Fisher's z-transform is applied. This correctly sensitizes the metric to changes near the critical boundaries.

$$\theta_{ij,t} = \tanh^{-1}(\rho_{ij,t}) = \frac{1}{2} \ln \left(\frac{1 + \rho_{ij,t}}{1 - \rho_{ij,t}} \right)$$

   These six values form the position vector $\boldsymbol{\theta}_t$.

4. **Velocity ($\mathbf{v}_t \in \mathbb{R}^6$) and Acceleration ($\mathbf{a}_t \in \mathbb{R}^6$):** The system's kinetics are derived from first and second-order finite differences of the position vector.

$$\mathbf{v}_t = \boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1} \quad ; \quad \mathbf{a}_t = \mathbf{v}_t - \mathbf{v}_{t-1} = \boldsymbol{\theta}_t - 2\boldsymbol{\theta}_{t-1} + \boldsymbol{\theta}_{t-2}$$

2.2 Global Dynamics (2 dimensions)

1. **Maximum Eigenvalue ($\lambda_{\mathbf{max},t}$):** This is extracted from the full 4×4 correlation matrix, $\boldsymbol{R}_t$. It quantifies the variance explained by the first principal component, serving as an excellent measure of system-wide coherence or "herding."

$$\lambda_{\max,t} = \max(\mathrm{eig}(\boldsymbol{R}_t))$$

2. **Eigenvalue Velocity ($\Delta\lambda_{\mathbf{max},t}$):** We also track its rate of change to measure how quickly system-wide contagion is building or receding.

$$\Delta\lambda_{\max,t} = \lambda_{\max,t} - \lambda_{\max,t-1}$$

2.3 State Vector Synthesis

The final step is to vertically concatenate the five component vectors into the single, comprehensive state vector:

$$\mathbf{s}_t = \begin{pmatrix} \boldsymbol{\theta}_t \\ \mathbf{v}_t \\ \mathbf{a}_t \\ \lambda_{\max,t} \\ \Delta\lambda_{\max,t} \end{pmatrix} \in \mathbb{R}^{20}$$

3 The Kinetic Stress Index (KSI)

3.1 Calculation via Mahalanobis Distance

The KSI is defined as the Mahalanobis distance between the current state vector $\mathbf{s}_t$ and its historical mean $\boldsymbol{\mu}_{t-1}$. This is a scale-invariant, geometry-aware distance that accounts for the historical variance and covariance of the system's components.

$$\mathrm{KSI}_t = \sqrt{(\mathbf{s}_t - \boldsymbol{\mu}_{t-1})^\top \boldsymbol{\Sigma}_{t-1}^{-1} (\mathbf{s}_t - \boldsymbol{\mu}_{t-1})}$$

The inputs are estimated over the lookback window $l = 252$:

$$\boldsymbol{\mu}_{t-1} = \frac{1}{l} \sum_{k=1}^{l} \mathbf{s}_{t-k}$$

$$\boldsymbol{\Sigma}_{t-1} = \mathrm{Cov}(\mathbf{s}_{t-l}, \ldots, \mathbf{s}_{t-1})$$

Because estimating a 20×20 covariance matrix from 252 data points is statistically challenging (the "curse of dimensionality"), a robust estimation technique is required. The code uses the

Graphical Inverse Shrinkage (GIS) method, a sophisticated non-parametric approach to find a well-conditioned inverse covariance matrix. For pedagogical purposes, this can be thought of as an advanced form of Ledoit-Wolf shrinkage, which regularizes the sample covariance matrix towards a structured target.

Intuition: The Mahalanobis Distance
Consider a 2D system with mean $\boldsymbol{\mu} = \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$ and covariance $\boldsymbol{\Sigma} = \left(\begin{smallmatrix} 9 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. This means the system typically varies much more along the x-axis than the y-axis.
Two points, $\mathbf{A} = \left(\begin{smallmatrix} 3 \\ 0 \end{smallmatrix}\right)$ and $\mathbf{B} = \left(\begin{smallmatrix} 0 \\ 3 \end{smallmatrix}\right)$, are both at a Euclidean distance of 3 from the center. However, their Mahalanobis distances reveal which is more "surprising":

$$D_M(\mathbf{A}) = \sqrt{\left(\begin{smallmatrix} 3 & 0 \end{smallmatrix}\right)\begin{pmatrix} 1/9 & 0 \\ 0 & 1 \end{pmatrix}\left(\begin{smallmatrix} 3 \\ 0 \end{smallmatrix}\right)} = \sqrt{1} = 1$$

$$D_M(\mathbf{B}) = \sqrt{\left(\begin{smallmatrix} 0 & 3 \end{smallmatrix}\right)\begin{pmatrix} 1/9 & 0 \\ 0 & 1 \end{pmatrix}\left(\begin{smallmatrix} 0 \\ 3 \end{smallmatrix}\right)} = \sqrt{9} = 3$$

Point **B** is three times more "stressful" or anomalous because it represents a 3-standard-deviation move along a direction of low historical variance, whereas point **A** is only a 1-standard-deviation move along a direction of high historical variance. The KSI applies this same logic in 20 dimensions.

3.2 Stress Decomposition and Attribution

A key strength of this framework is that a high KSI value can be diagnosed. By performing an eigendecomposition on the covariance matrix ($\boldsymbol{\Sigma}_{t-1} = \boldsymbol{V}\boldsymbol{L}\boldsymbol{V}^\top$, where $\boldsymbol{L} = \operatorname{diag}(\ell_1, \ldots, \ell_{20})$), we can express the total squared KSI as a sum of 20 independent contributions.

$$\mathrm{KSI}_t^2 = (\mathbf{s}_t - \boldsymbol{\mu}_{t-1})^\top (\boldsymbol{V}\boldsymbol{L}^{-1}\boldsymbol{V}^\top)(\mathbf{s}_t - \boldsymbol{\mu}_{t-1}) \tag{1}$$

$$= \sum_{j=1}^{20} \frac{y_j^2}{\ell_j}, \quad \text{where } \mathbf{y} = \boldsymbol{V}^\top(\mathbf{s}_t - \boldsymbol{\mu}_{t-1}) \tag{2}$$

The term $C_j = y_j^2/\ell_j$ is the contribution to total squared stress from the j-th eigenmode. This represents the squared deviation along mode j, normalized by the typical variance of that mode.

To make this interpretable, we attribute the stress from each mode back to the original 20 components. The squared loading of component i on eigenvector j, denoted v_{ij}^2, tells us what fraction of mode j is described by component i. The total contribution of component i, Stress_i, is the sum of its contributions from all 20 modes:

$$\text{Stress}_i = \sum_{j=1}^{20} C_j \cdot v_{ij}^2$$

This allows us to calculate the percentage of total stress coming from, for example, the acceleration of a specific correlation pair.

4 Interpreting the Code and its Visualizations

The Python scripts produce a suite of outputs, each with a specific purpose rooted in the mathematics above.

4.1 Main Analyzer Output

- **KSI Time Series Plot:** This is the primary output, showing KSI_t over time. High values indicate that the system's 20D state vector is in a statistically anomalous region relative to its one-year history.
- **Console Report & Peak Analysis:** The script identifies the highest, most separated KSI peaks. For each peak, it uses the stress attribution method to calculate the percentage contribution from four broad categories, which are defined in the '$_c lassify_c omponents$'$method:$

  - **Intra-Kinetics:** Stress from pairs within the same group (e.g., Policy-Policy).
  - **Cross-Other (Position/Velocity):** Stress from changing correlation levels between groups.
  - **Cross-Accel:** Stress from the *acceleration* of correlations between groups. This is a key indicator of sudden, forceful cross-contamination.
  - **Global Eigen:** Stress from the system-wide herding measures.

Visual Intelligence Plots:

- *Regime Fingerprints:* A dashboard showing the long-term evolution of the system's correlation structure, phase space, and kinetic energy.
- *Event Gallery & Detailed Event Analysis:* These use network diagrams to visualize the correlation matrix $\boldsymbol{R}_t$ on specific high-stress dates. The pie chart in the detailed analysis is a direct visualization of the stress attribution percentages described above.

4.2 Follow-up: Rolling Stress Attribution Plot

This crucial visualization, generated by the second script, provides a temporal view of the stress attribution.

- **Methodology:** It first computes the daily stress attribution percentages for the four categories above. It then applies a rolling average (e.g., 60-day window) to this data to smooth out daily noise and reveal the underlying **stress regime**.
- **Top Panel (Stacked Area Chart):** This shows the evolution of the smoothed stress contributions over time. It answers the question: "Over the last quarter, what has been the primary source of market tension?" A shift from a yellow-dominant (Intra-Kinetics) regime to a red-dominant (Cross-Accel) regime signifies a dangerous change from contained stress to systemic contagion.
- **Bottom Panel (Dominance Ratio):** This plot makes the regime shift explicit by charting the ratio:

  $$\text{Dominance Ratio}_t = \frac{\text{Cross-Contamination Stress}_t}{\text{Total Stress}_t}$$

  When this ratio is above 0.5, it signals that the tension between short-term and long-term fears has become the dominant driver of market risk, a potentially unstable condition.