SUMMARY OF MY RESEARCH

NAM TRANG

My research interest in pure mathematics is in set theory and set-theoretic topology. My current research focuses on studying models of various fragments of determinacy for games of infinite length on natural numbers or reals, studying the connections between inner models, structures of sets of reals (like those of determined sets in Polish spaces and quotient spaces of Polish spaces by various equivalence relations), hybrid structures (such as HOD[1] of determinacy models), forcing, and strong combinatorial principles. I'm also interested in applications of strong forcing axioms such as PFA and their connection with other combinatorial principles in set theory. As mentioned, my other interest in pure mathematics is in set-theoretic topology, in particular the study of Ramsey-like spaces on functions on infinite ordinals. Some of the techniques mentioned above are used in combination with combinatorial techniques developed specifically to understand and solve various fundamental topological and combinatorial problems in Ramsey-like spaces of functions on various infinite cardinals. This work has earned me two regular NSF grants and currently an NSF CAREER grant. I describe this work in Section I.

My research interest in applied mathematics includes various aspects of mathematical finance and applied statistics, and in deep learning as applied to computer vision. My work in computer vision, as part of a team of industry expert, is being supported by a DARPA grant. I describe some ongoing work in this area in Section II.

I. SET THEORY AND ITS APPLICATIONS

1. Inner models and determinacy

The research described below mostly belongs to an area of set theory called *descriptive inner model theory* (DIMT). DIMT is an emerging field in set theory that explores deep connections between *descriptive set theory* (DST) and *inner model theory* (IMT). DST studies a certain class of well-behaved subsets of the reals and of Polish spaces (e.g. Borel sets, analytic sets) and has its roots in classical analysis, through work of Baire, Borel, Lebesgue, Lusin, Suslin and others. One way a collection Γ of subsets of a Polish space can be well-behaved is that they satisfy various regularity properties, e.g. they have the Baire property, every uncountable set in Γ contains a perfect subset, every set in Γ is Lebesgue measurable. A cornerstone in the history of the subject is the discovery of the Axiom of Determinacy (AD) by Mycielsky and Steinhaus in 1962. AD states that every infinite-length, two-person game of perfect information where players take turns play integers is determined, i.e. one of the players has a winning strategy. If every set in $\Gamma \subseteq \mathcal{P}(\mathbb{R})$ is determined, then they have all the regularity properties listed above (and more), and hence very well-behaved. AD contradicts the axiom of choice as the latter implies the existence of very irregular sets like the Vitali set; however, inside a universe of ZFC, there may be many interesting sub-universes (models) that satisfy AD, for instance, $L(\mathbb{R})$ the minimal transitive class model of ZF that contains the reals may satisfy AD. One important and fruitful branch of descriptive set theory studies structure theory of models of AD. IMT forms one of the core subjects in modern set theory; its main objective is study "canonical" models of various extensions of ZFC, called *large cardinal axioms* (or simply large cardinals) and construct such models under various circumstances (e.g. see question (2) below). The large cardinal axioms form a linear hierarchy of axioms (in terms of consistency strength) extending ZFC and every known, natural axiom in mathematics/set theory is decided by one such axiom.[2] The first "canonical model" of large cardinals is the *Gödel's constructible universe* L, the minimal model of ZFC. It is well-known that L cannot admit "very large" large cardinals; *the Gödel's inner model program*, a major program in inner model theory, aims to construct and analyze L-like models that can accommodate larger large cardinals under various hypotheses. Benchmark properties that help determine the canonicity of these models include *the Generalized Continuum Hypothesis* (GCH), Jensen's $\square$-principles (see Question (1), more details later).

DIMT uses tools from both DST and IMT to study and deepen the connection between canonical models of large cardinals and canonical models of AD. One of the first significant developments in DIMT comes in the 1980's with works of Martin, Steel, Woodin and others; their work, for instance, shows that one can construct models of

[1] HOD stands for the class of "Hereditarily Ordinal Definable" sets, see [11] for a definition.

[2] By Gödel's incompleteness theorem, given any axiom (A) extending Peano Arithmetic, one can construct a sentence, albeit unnatural, which is not decidable by (A).

AD (e.g. they showed AD holds in $L(\mathbb{R})$) assuming large cardinal axioms (those that involve the crucial notion of Woodin cardinals) and conversely, one can recover models of large cardinal axioms from models of AD. The key to uncover these connections is to analyze structure theory of models of AD. Much of current research in DIMT and my research go along this line (e.g. see Question (1) below). One of the main goals of this analysis is to understand well enough canonical models of AD and those of large cardinals to compute consistency strength of strong theories such as the Proper Forcing Axiom (PFA). The conjecture that "PFA has the exact consistency strength as that of a supercompact cardinal" is one of the most longstanding and arguably important open problems in set theory. Techniques recently developed in DIMT and to some extent from the research described here also enable us to make significant progress in calibrating the consistency strength lower-bound for PFA (Question (2)).

We now describe these connections in more technical details in the next couple of paragraphs.[3] One way of formalizing this connection is through *the Mouse Set Conjecture* (MSC), which states that, assuming the AD or a more technical version of it (AD^+), then whenever a real x is ordinal definable from a real y, then x belongs to a canonical model of large cardinal (mouse) over y. MSC conjectures that the most complicated form of definability can be captured by canonical structures of large cardinals. An early instance of this is a well-known theorem of Shoenfield that every Δ^1_2 real is in the Gödel's constructible universe L. Another instance is a theorem of W.H. Woodin's that in the minimal class model containing all the reals, $L(\mathbb{R})$, if AD holds, then MSC holds. However, the full MSC is open and is one of the main open problems in DIMT.[4] Instances of MSC have been proved in determinacy models constructed by their sets of reals (much larger than $L(\mathbb{R})$) and these proofs typically obtain canonical models of large cardinals (mice) that capture the relevant ordinal definable real by analyzing HOD of the determinacy models. Hence, the key link between these two kinds of structures (models of large cardinals and models of determinacy) is the HOD of the determinacy models. A central notion in the proof of MSC and the analysis of HOD is the notion of *hod mice* (developed by G. Sargsyan, cf. [18], which built on and generalized earlier unpublished work of H.W. Woodin), which features heavily in my research described here. Hod mice are a type of models constructed from an extender sequence and a sequence of iteration strategies of its own initial segments. The extender sequence allows hod mice to satisfy some large cardinal theory and the strategies allow them to generate models of determinacy. Unlike pure extender models, there are many basic structural questions that are still open for hod mice (to be discussed later).

Another way of exploring the determinacy/large cardinal connection is via the Core Model Induction (CMI), which draws strength from natural theories such as PFA to inductively construct canonical models of determinacy and those of large cardinals in a locked-step process by combining core model techniques (for constructing the core model K) with descriptive set theory, in particular the scales analysis in $L(\mathbb{R})$ and its generalizations. CMI is the only known systematic method for computing lower-bound consistency strength of strong theories extending ZFC and it is hoped that it will allow one to compute the exact strength of important theories such as PFA. CMI is another central theme of my research. Much work has been done the last several years in developing techniques for CMI and it's only recently realized that constructing hod mice in non-AD^+ contexts (e.g. in the context of PFA, in contrast to Sargsyan's constructions) is a key step in CMI.

A large part of my research described here is devoted to studying structural properties of hod mice, developing methods for constructing hod mice in non-AD^+ contexts, and applying these constructions in the core model induction. One of the main goals is to construct determinacy models of "$\mathsf{AD}_{\mathbb{R}} + \Theta$ is regular", "$\mathsf{AD}^+ +$ the largest Suslin cardinal is on the Solovay sequence" (LSA) and beyond, from various theories such as PFA. These lower-bounds are beyond the reach of pure core model methods.

There are many problems in DIMT that I've been interested in and working on but my hope is to make progress toward answering two fundamental questions in the area (to be discussed in details later):

(1) Is HOD of a determinacy model fine-structural (e.g. do the $\mathsf{GCH}, \square$ hold in HOD)? What large cardinals can HOD accommodate?

(2) What is the consistency strength of PFA?

The three main aspects of my research are as follows: (a) connections between inner models, hybrid structures, and canonical sets of reals; (b) applications of the structure theory of the three hierarchies in (a) and their connections; (c) strong combinatorial principles, determinacy, and large cardinals through the lens of forcing. Many of the important problems in these areas are either (indirectly) related to or (directly) elaborated from problems (1) and

[3]Non-logicians may want to simply skim this through.

[4]It is widely believed that a more fundamental property called *generation of pointclasses* or *hod pair capturing*. Current research focuses on proving (instances of) this property.

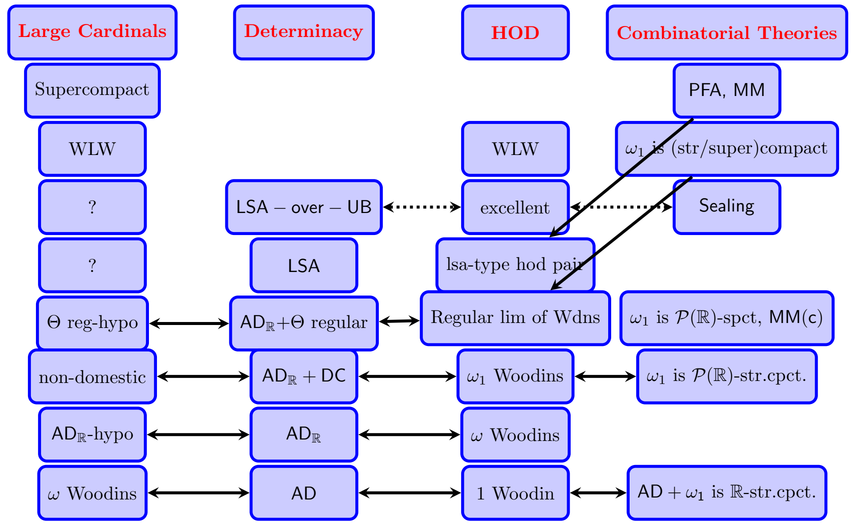

Figure 1: Consistency strength of various theories

(2) above and hence quite ambitious. However, I think these problems are worth pursuing as they are important for advancing the field. I have obtained results related to most of them in the past (cf. [23], [19], [15], [26], [25], [20], [21] etc.). See my website (http://math.unt.edu/∼ntrang) for full my research publications. Many of the connections between consistency strength of theories are summarized by Figure 2; several of these connections have been established by the work described above. Much of the work in this area that I plan to pursue in the next several years is to develop tools (in the basic theory of HOD and the core model induction) to approach these fundamental problems.

I summarize some of the major theorems proved recently regarding the topics discussed above. The first one is the main theorem of [25].

Theorem 1.1 (Trang)**.** *Assume* PFA*. Then there is an inner model M such that $M \vDash$ "$\mathsf{AD}_{\mathbb{R}} + \Theta$ is regular".*

G. Sargsyan and I, in [19][5], significantly improve the above theorem. We show.

Theorem 1.2. *Assume* PFA*. Then there is an inner model M such that $M \vDash$* LSA*.*

In [21], Sargsyan and I have computed the exact consistency strength of Woodin's "Γ_∞ Sealing" principle and identified Sealing of Γ_∞, the universally Baire sets, as an obstruction to inner model theory (in the sense discussed there). Some of the current work I'm pursuing is developing methods for overcoming such obstructions. One instance of this is an ongoing work that constructs a non-tame hod mouse from PFA. The main point of this construction is that we construct a canonical subset of Γ_∞ and show it generates a strategy for a non-tame mouse. This is significant because the existence of a non-tame hod mouse is stronger, consistency-wise, than "Γ_∞ Sealing", which in turns is stronger than LSA.

Finally, in [1], my co-authors and I prove the following theorem, resolving part of Problem 12 in [27].

Theorem 1.3. *Assume* CH *and there is an ω_1-dense ideal on ω_1. Then there is an inner model M such that $M \vDash$ "$\mathsf{AD}_{\mathbb{R}} + \Theta$ is regular".*[6]

2. Set-theoretic topology

Ramsey famously showed that the set of natural numbers, ω, satisfies the finite partition relations $\omega \to (\omega)^k_2$ for each $k < \omega$, i.e. for any (coloring) function $\Phi : [\omega]^k \to 2$, there is an infinite homogeneous set for Φ, i.e. an infinite set $C \subset \omega$ such that $\Phi \restriction [C]^k$ is constant. The infinite exponent partition relation $\omega \to (\omega)^\omega_2$ (also called the Ramsey property for all partitions) is a natural generalization that is not compatible with the axiom of choice.

[5]This manuscript is to be published in the Lecture Notes in Logic series, by the Cambridge University press.

[6][1] has been accepted by the Journal of the American Mathematical Society. This is joint work with D. Adolf, G. Sargsyan, T. Wilson, M. Zeman.

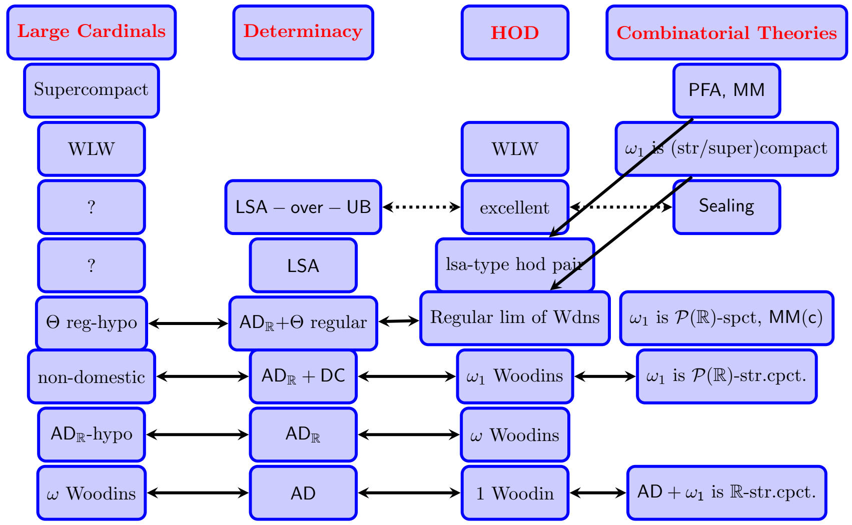

Figure 2: Consistency strength of various theories

However, simple definable partitions such as Borel or analytic partitions always satisfy the Ramsey property (see [9] and [22]).

Mathias, in his celebrated work [13], studied the almost everywhere behavior of functions on the Ramsey space $[\omega]^\omega$ of functions from ω to ω such as when every function $\Phi : [\omega]^\omega \to \mathbb{R}$ is Ramsey almost everywhere continuous or every relation. $R \subseteq [\omega]^\omega \times \mathbb{R}$ has a Ramsey almost everywhere uniformization.

In past and ongoing joint work with W. Chan and S. Jackson, which has resulted in a series of papers (see for example [6, 2, 3]), we attempt to study similar properties for functions on the Ramsey-like spaces $[\kappa]^\kappa$ (or $[\kappa]^\epsilon$) for infinite cardinals $\kappa \geq \epsilon$ and apply these properties to various applications in set theory and set-theoretic topology, in particular the study of cardinalities without the axiom of choice (or with limited choice). This research program not only is fruitful in attacking fundamental problems about these function spaces but also has applications in various problems in studying combinatorial objects in the context of AD. For instance, [4] characterizes the existence of maximal almost disjoint families on every cardinal κ in natural models of AD^+.

Unsurprisingly, (almost everywhere) continuity and monotonicity of functions on these spaces are false in ZFC. However, under the Ramsey property, for instance $\kappa \to (\kappa)^\kappa_2$, we can prove monotonicity for functions on the space $[\kappa]^\kappa$. A set $C \subseteq \kappa$ is club in κ if it is closed (under the topology induced by the well-order of the ordinals below κ) and unbounded in κ. ON denotes the class of ordinals. The cofinality of a cardinal κ is the length of the shortest increasing, unbounded, and continuous sequence in κ.

Theorem 2.1 (Chan-Jackson-Trang, [5]). *Suppose $\kappa \to (\kappa)^\kappa_2$. For any function $\Phi : [\kappa]^\kappa \to ON$, there is C club in κ so that for all $f, g \in [C]^\kappa$[7], if for all $\alpha < \kappa, f(\alpha) \leq g(\alpha)$, then $\Phi(f) \leq \Phi(g)$.*

See [5] for a much more detailed analysis and refinements of the above theorem as well as several theorems on (almost everywhere) continuity of functions of Ramsey-like spaces, which is a more fundamental property and is used (along with other combinatorial techniques we developed) to establish monotonicity results like the above. The following is a representative theorem about continuity on these spaces.

Theorem 2.2 (Chan-Jackson-Trang, [5]). *Suppose κ is a cardinal, $\epsilon < \kappa$ is a limit ordinal with cofinality $cof(\epsilon) = \omega$, and $\kappa \to (\kappa)^{\epsilon\times\epsilon}_2$ holds. For any function $\Phi : [\kappa]^\epsilon \to ON$, there is a club $C \subseteq \kappa$ and a $\delta <$ so that for all $f, g \in [C]^\epsilon \to ON$, if $f \restriction \delta = g \restriction \delta$, and $sup(f) = sup(g)$, then $\Phi(f) = \Phi(g)$.*

II. APPLIED MATH

3. Mathematical Finance, Graphs and Matrices

I am interested in various aspects of mathematical finance. In a joint ongoing project with M. Foreman and D. Brownstone, we are interested in: empirically verifying how accurate asset pricing theories such as the Capital Asset Pricing Model (CAPM), the Arbitrage Pricing Theory (APT) and its variations are, and in designing novel

[7]We tacitly assume all functions are increasing.

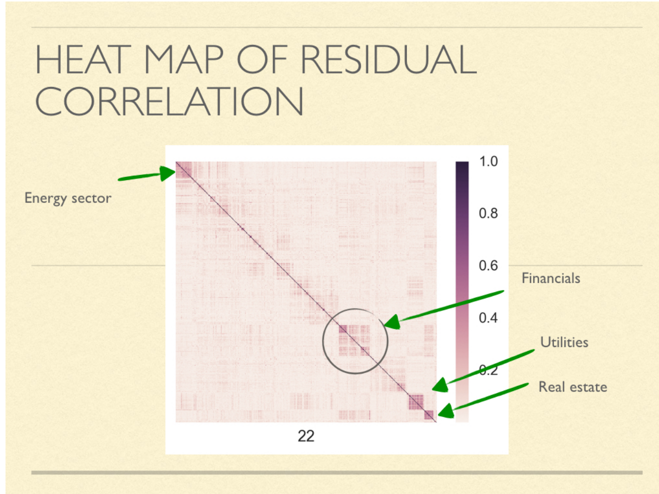

Figure 3: Correlated Sectors

methodologies for predicting stock prices using short-time series of stock data (e.g. stock data of the previous 6 months).

The methodology is as follows (this is the first approximation). Suppose we have N independent and identically distributed (iid) random variables $X_1, \dots, X_N$. X_i represents the data on the i-th stock. In practice, we get a time series of stock returns and these series are (approximately) given by the X_i's. Given a threshold $0 < \delta < 1$, we randomly sample (without replacements) within the set $\mathcal{X} = \{X_1, \dots, X_N\}$ until we capture $1-\delta$ of the variance of the stocks.[8] In particular, if $\mathcal{X}$ represents the stocks in the S&P 500 and say $\delta = 0.1$, we use the random sampling method to obtain a set $\mathcal{P} \subset \mathcal{X}$ such that when regress the S&P 500 stocks on $\mathcal{P}$, we get a $R^2 \geq 1 - \delta = 0.9$. We repeat this process to get averages and confidence intervals. The advantage of this method is that we reach the regression threshold with a relatively small number of stocks (e.g. 15) for the S&P 500; this is more feasible than trying to compute the eigenvalues of a 500×500 matrix. However, this methodology is flawed. Figure 3 represents the residuals of other stocks when projected onto the hyperplane spanned by $\mathcal{P}$.

In the above figure, the method did not pick up various sectors that are highly correlated; these correlations are meaningful economically. When we add these stocks into our portfolio, the number of stocks is close to 30 (on average). This suggests better algorithms for constructing random portfolios such as "cluster sampling" ala the census bureau. The high correlations means that our random selections have not touched these sectors – we should artificially pick from those sectors and revise our assumption about the uniform distribution of coupons. We haven't done this study yet. We also hope to apply the same methodology (and its improvements) to other collections of stock index, like the Russell 2000 etc. The hope is to increase the robustness of our methods (i.e. the bigger the stock universe is, the better it approximates the "whole economy") and to empirically verify whether theories like APT hold up when pricing stocks in various sectors of the economy.

The following conjecture is a more general statement that captures our attempts above. The idea is in order to capture enough of the total variance, we need enough "dimensions", i.e. the hyperplane generated by the stocks we sample must be "close enough" to all the stocks in the original collection. Below, for any two random variables X, Y, let $Cov(X,Y) = \langle X, Y \rangle$ be the covariance of the X, Y; let $Var(X) = ||X||^2$ be the variance of X.[9] For a hyperplane $\mathcal{P}$ and a random variable X, we let $X^{\mathcal{P}}$ be the projection of X onto $\mathcal{P}$. In the following, we let $\mathcal{C} = (\langle X_i, X_j \rangle)_{1 \leq i,j \leq n}$ be the covariance matrix of $\{X_1, \dots, X_N\}$; by the Singular Valued Decomposition (SVD) theorem, we get a diagonal matrix matrix $\mathcal{D} = diag(\lambda_1^2, \dots, \lambda_N^2)$ and a unitary matrix U such that $\mathcal{D} = U^T \mathcal{C} U$. We also have an orthogonal set of vectors $\{Y_1, \dots, Y_N\}$ with the property that $\langle Y_i, Y_i \rangle = \lambda_i^2$.

Conjecture 3.1 (Unindexed version)**.** *For any $0 < \lambda < min_{i \leq N} \lambda_i^2$, for any $I \subseteq \{1, \dots, N\}$, for any $\delta = \sum_{i \in I} \lambda_i^2 + \lambda$, there is $\epsilon(\delta)$ and $H \subseteq \{X_1, \dots, X_N\}$ such that letting $\mathcal{P}$ be the hyperplane generated by H, the following are equivalent:*

[8] This is a variation of the Coupon Collection Problem.

[9] It is not a coincidence that we denote Cov and Var as inner product and norm. The are indeed the $\mathcal{L}_2$ inner product and norm of real-valued functions on the sample space. We treat a random variable as a vector in the $\mathcal{L}_2$ space.

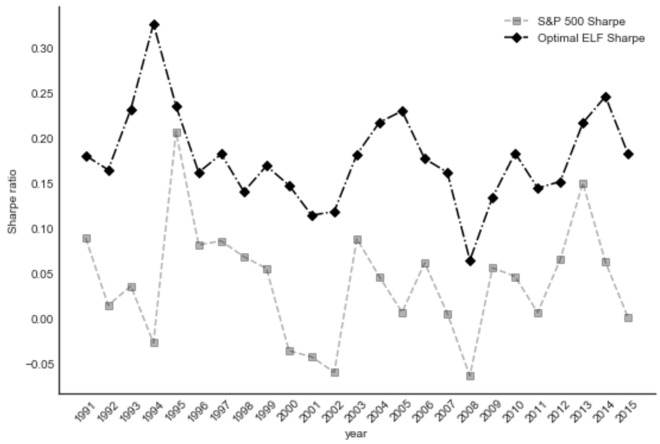

Figure 4: Sharpe ratio comparisons

- $$\frac{\sum_{i=1}^{N}||X_i - X_i^{\mathcal{P}}||^2}{\sum_{i=1}^{N}||X_i||^2} < \delta$$

- *for some* $\vec{Y} \subseteq \{Y_j : j \leq N\}$

  1. $$\sum\nolimits_{Y_j \in \vec{Y}} \frac{||Y_j - Y_j^{\mathcal{P}}||^2}{||Y_j||^2} < \epsilon(\delta)$$

     and

  2. $$\frac{\sum_{Y_j \notin \vec{Y}}||Y_j||^2}{\sum_{i=1}^{N}||Y_j||^2} < \delta$$

One can also state the indexed version of the above conjecture. We have reasons to believe the conjectures are true in cases of interest, namely when a small number of eigenvalues of $\mathcal{C}$ are much larger than the rest. We hope that the conjectures (and its generalizations) will shed light on the relationships between the sampling method (better yet, its improvements) and pricing models. In particular, we hope to show that no reasonable number of factors (e.g. Ross [17], Fama-French [7]) can capture (most of) the economy (in terms of returns of stocks and bonds).

The paper [8] is one along this direction. The Capital Asset Pricing Model (CAPM) is a ubiquitous tool for financial applications, from asset management to corporate decision making. It is simply stated, has elegant consequences and has easily applicable corollaries. Unfortunately, due to inherent estimation difficulties it is difficult to check directly. This paper describes a mathematical technique, the Census Taker Method (CTM) which bypasses these estimation difficulties and makes conservative estimates of efficient frontiers. A direct application is to show that the $S\&P$ 500 does not realize returns anywhere close to its efficient frontier (see Figure 4, where it is shown that the Sharpe ratio of the S&P 500 is nowhere closed to that of estimated long efficient frontiers using CTM, hence the Sharpe ratio of the S&P 500 is very far from actual efficient frontiers).[10] Thus a central corollary of the CAPM, at least as used in practice, is empirically false.

In ongoing joint-work with Prof. K. Song at UNT, we are interested in the following problem.

Problem 3.2. *Given an* $n \times n$*, symmetric matrix* $A = (a_{i,j})$*. Find a* λ *such that*

1. $A + \lambda I_n$ *is diagonally dominant.*[11]
2. $||A + \lambda I_n||_\infty - ||A||_\infty$ *is minimal.*[12]

[10] The figure is the result of running the CTM on S&P 500 stock data.
[11] I_n is the $n \times n$ identity matrix. A matrix $(a_{i,j})$ is diagonally dominant if for every i, $|a_{i,i}| > \sum_{j \neq i} |a_{i,j}|$.
[12] We can replace the ∞-norm by other matrix norms. It is not too important what norm we use.

3. *$(A+\lambda I_n)^{-1}$ exists, and $(A+\lambda I_n)^{-1}, \mathcal{D}((A+\lambda I_n)^{-1})$can be computed or approximated efficiently in a stable fashion.*[13]

This general problem is of interest in many applications, including in machine learning and statistics. The matrix A is typically the Hessian matrix of some appropriate function or the covariance matrix that come from data that arise in applications. In many cases, it is numerically unstable to compute A^{-1}. We may, instead, want to "perturb" A and define a matrix B (in our case $B = A + \lambda I_n$) that is slightly different from A (see requirement (2)) and yet, we can compute or approximate B^{-1} efficiently and the computations are numerically stable.

In many applications, it is often times okay to use B^{-1} as a substitute for A^{-1}. For instance, in optimization or machine learning, one can execute the line search algorithm using the Newton's method applied to a given function f: at step k, the k-th search direction is $x_k = x_{k-1} + \epsilon_k p_k$, where $p_k = H_{k-1}^{-1}\nabla f(x_{k-1})$ and H_{k-1} is the Hessian of f at the point x_{k-1} and ϵ_k is the step size. For convex f, Newton's method guarantees convergence, but it usually is not feasible to compute H^{-1} at each step. By replacing H_k by B_k that satisfies certain conditions (like B_k^{-1} can be computed efficiently and the condition numbers of the B_k's are uniformly bounded, cf [14]), we can still guarantee convergence and efficiency in execution the line search.

The standard algorithm for computing B^{-1} has $O(n^3)$ complexity. Much more efficient algorithms have been designed for various special types of matrices (banded matrices, sparse matrices etc.), see [10]. Our approach to resolving the above problem is via computing the Neumann expansion of $B = A + \lambda I_n$. (cf. [24]). We can show that if λ is chosen such that B is diagonally dominant and positive definite, then we can approximate B^{-1} using the first p many terms of the Neumann expansion of B^{-1} for a relatively small p. p depends on various characteristics of A but our numerical experiments show that in most cases, $p << n$.

In many applications, one often does not care about B^{-1}, but instead computing/approximating $\mathcal{D}(B^{-1})$ becomes the focus. We can achieve this via the *method of probing vectors* (and very efficiently in the case where A is sufficiently sparse, see [24] for a detailed complexity analysis of this method).[14] We just highlight one interesting aspect of this method and how it connects computational linear algebra with graph theory. We define the adjacency graph $G(B) = (V, E)$ associated with the symmetric matrix $B = (b_{i,j})$ as follows: *the vertices $V = \{b_i : i < n\}$, where b_i is the i-th column of B, and $(a_i, a_j) \in E$ if and only if $j \neq i$ and $b_{i,j} \neq 0$.* A key point in the probing vectors method is in coming up with algorithms for coloring the graph $G(B)$. An efficient algorithm using a small number $p << n$ of colors will result in an efficient approximation of $\mathcal{D}(B^{-1})$, see [24]. Our experiments indicate that this method is very numerically stable and has worked very well for certain classes of matrices A (e.g. sparse matrices). Furthermore, I have proven relevant theorems that give bounds on the *chromatic number* of $G(B)$ for a class of matrices B that we are interested in (e.g. for B which are block diagonal with a small number of non-zero entries off of the diagonal blocks, then the chromatic number is bounded by the size of the largest block, or more generally, if B is symmetric with only $2k$ many nonzero entries, then the chromatic number is $O(\sqrt{k})$).

An ongoing work on matrix factorization concerns the following conjecture on the exact cancellations of Cholesky factorizations of symmetric positive definite matrices (SPD). In the following, let $A = LL^T$ be the Cholesky factorization of $A = (a_{ij})$ for $1 \leq i, j \leq n$; here $L = (l_{ij})$ is lower-triangular. We say that an (i, j) element is an *exact cancellation* if $a_{ij} \neq 0$ but $l_{ij} = 0$. We say (i, j) is a *fill-in* if $a_{ij} = 0$ but $l_{ij} \neq 0$. These concepts have received considerable attention from applied mathematicians and computer scientists (see [12, 16]) as they are useful in constructing efficient factorizations of sparse SPDs. In particular, a graph structure (called elimination graph) has been used to prove that assuming the Cholesky factorization of an irreducible matrix A (or its permutation) has no exact cancellations, then some permutation of A produces a Cholesky factorization with no fill-ins. However, the question of whether Cholesky factorizations of some permutation of A will suffer no exact cancellations remains open. In practice, since matrices have entries that are floating point numbers, having no exact cancellations almost always occur. However, from the mathematical perspective, the following conjecture is very interesting.

Conjecture 3.3. *Given an SPD A, there is a permutation $A' = P^T AP$[15] of A such that the Cholesky factorization of A' has no exact cancellations.*

The conjecture has been proven to be true for matrices up to 5×5 (by continuing work of mine and an undergraduate student). With Joseph Prein (a graduate student working under my supervision), we have shown the following general theorem. The first Schur's complement of and $n \times n$ SPD A is defined to be the $(n-1) \times (n-1)$ matrix $S^{(1)} = (a_{ij})_{i,j\geq 2} - \frac{1}{a_{11}}(a_{k1})_{k\geq 2}(a_{k1})^T_{k\geq 2}$.

[13] $\mathcal{D}(A)$ is the matrix whose entries along the diagonal are the diagonal entries of A and whose off-diagonal entries are 0.

[14] In practice, if our matrix A is not sparse, but has many non-zero entries that are very close to 0, we threshold A, by considering considering the matrix A_ϵ for some small ϵ, where the (i, j)-entry of A_ϵ is $a_{i,j}$ if $|a_{i,j}| \geq \epsilon$ and is 0 otherwise.

[15] P is called a permutation matrix.

Theorem 3.4. *Given an irreducible SPD A. There is a permutation $A' = P^T AP$ of A such that there is a column in $S^{(1)}$ that has no exact cancellations.*

Since the Schur's complement is used in the inductive definition of Cholesky factorization that produces the columns of L, we hope this will be important in settling the full conjecture. My undergraduate student is exploring various ideas related to this.

In another joint work with students, we have finished a paper (with a graduate student and two undergrads) on the topic of games on a certain types of graphs, called partition graphs. The paper deals with the games of Cops and Robber on tunnel graphs. We prove several theorems that relate the cop number of the game on a given graph G and the corresponding game on the n-tunnel graph G_n (basically the cop number can increase by at most one). The paper, entitled "On the cop number of subdivisions of graphs", has been submitted to the Journal of Combinatorics.

Many related problems in these areas are suitable as research topics for undergraduate students and early graduate students. For instance, I believe motivated undergraduates can undertake projects similar to the above (on various types of games on graphs). Sparse matrix approximations is another area where graph-theoretic techniques can be applied (approximation of inverse of sparse matrices, efficient approximations of solutions of linear equations of sparse matrices etc.) and undergraduate students can participate in research projects.

4. Deep Learning and Computer Vision

Regarding industry work, I'm part of a team of industry experts working on bio-inspired, adaptive artificial intelligence (AI) systems that are suitable for many types of applications, including image processing, image recognition, feature extractions etc. Our company is working on a DARPA funded project that uses our technology to solve certain problems/tasks DARPA is interested in.

The research project, funded by DARPA, is on designing and building an in-pixel intelligent system as described Figure 5. The system emulates the saccadic eye movement (hence bio-inspired) in processing visual information. It consists of the design of a chip that processes images with resolution $1k \times 1k$, produces a sparsification ($10\times$-reduction) of the image (called the P, F, and the L features), and doing all of this with a constraint in power consumption (at most 250 mWatts per image). The sparse images retain enough visual information so that the back-end (an image detection/recognition system) can be used to perform various tasks on them (like training on these sparse images and feeding the information back into the on-chip memory and the focal plane). The whole system is a close-loop system that achieves high-accuracy (in terms of its primary task, namely recognition, detection and tracking) and at the same time is low-power and fast (at least 100 frames per second). See Figure 6 for a block diagram of the system. The technology used here has potential for commercialization and many other applications in computer vision. This is another key research direction that I plan to pursue in the next several years. The main goal is to refine and improve the above technology and build around that core technology so that it's easy to adapt to various applications.

One major potential application in computer vision is to design "compact" neural network models that offer comparable performance to "deep" neural networks but take much less time to train. For instance, one of the sparsification techniques mentioned above have the potential to help reduce the size of the networks for certain computer vision tasks (like image detection). The technology allows us to isolate "good" features as opposed to obtaining them via training convolutional networks with many layers and each layer has many filters (i.e., features) to train. This is the key point of my current research in this direction. If successful, we hope to obtain convolutional neural networks that consist of very few convolutional layers (e.g., 1 to 2 layers) in contrast to current state-of-the-art neural networks (e.g., the YOLO models and its variations) that may contain dozens of such layers.

This is another area that undergraduates can take part in research. For instance, I am mentoring an intern, who is a senior undergraduate at UC Irvine, at the company I'm working with on some work related to computer vision and deep learning. I would envision many projects that undergraduates can participate in (e.g., build convolutional neural networks that perform certain tasks in computer vision). Students will learn very interesting mathematical ideas (like convolutions, back-propagation, logistic regression, and many more that go into the building blocks of a neural network) and practical programming skills. These are absolutely important skills, I believe, that a STEM student (in most disciplines) should acquire.

References

[1] Dominik Adolf, Grigor Sargsyan, Nam Trang, Trevor Wilson, and Martin Zeman. Ideals and strong axioms of determinacy. *accepted to the Journal of the American Mathematical Society*, 2022. available at

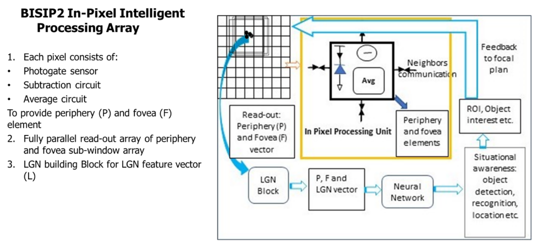

Figure 5: A high level diagram of the In-pixel intelligent processing system

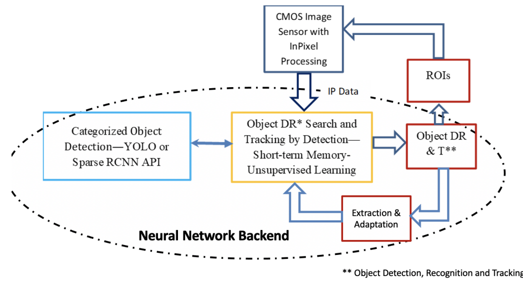

Figure 6: Block diagram of the system

https://math.unt.edu/∼ntrang.

[2] William Chan, Stephen Jackson, and Nam Trang. More definable combinatorics around the first and second uncountable cardinal. *submitted, available at https://math.unt.edu/∼ntrang*, 2020.

[3] William Chan, Stephen Jackson, and Nam Trang. Countable length everywhere club uniformization. *to appear on the Journal of Symbolic Logic, available at https://math.unt.edu/∼ntrang*, 2021.

[4] William Chan, Stephen Jackson, and Nam Trang. Almost disjoint families under determinacy. 2022. submitted, available at https://math.unt.edu/∼ntrang.

[5] William Chan, Stephen Jackson, and Nam Trang. Almost everywhere behavior of functions according to partition measures. *submitted, available at https://math.unt.edu/∼ntrang*, 2022.

[6] William Chan, Stephen Jackson, and Nam Trang. The size of the class of countable sequences of ordinals. *Transactions of the American Mathematical Society*, 375(03):1725–1743, 2022.

[7] Eugene F Fama and Kenneth R French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56, 1993.

[8] Matt Foreman, Nam Trang, and David Brownstone. An empirical refutation of CAPM. *In preparation*, 2020-2021.

[9] Fred Galvin and Karel Prikry. Borel sets and ramsey's theorem1. *The Journal of Symbolic Logic*, 38(2):193–198, 1973.

[10] Gene H Golub and Charles F Van Loan. Matrix computations johns hopkins university press. *Baltimore and London*, 1996.

[11] Thomas Jech. *Set theory*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2003. The third millennium edition, revised and expanded.

[12] Joseph WH Liu. The role of elimination trees in sparse factorization. *SIAM journal on matrix analysis and applications*, 11(1):134–172, 1990.

[13] Adrian RD Mathias. Happy families. *Annals of Mathematical logic*, 12(1):59–111, 1977.

[14] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[15] D. Rodríguez and N. Trang. $L(\mathbb{R}, \mu)$ is unique. *Advances in Mathematics*, 2018.

[16] Donald J Rose. A graph-theoretic study of the numerical solution of sparse positive definite systems of linear equations. In *Graph theory and computing*, pages 183–217. Elsevier, 1972.

[17] Stephen A Ross. *Neoclassical finance*. Princeton University Press, 2009.

[18] G. Sargsyan. *Hod mice and the mouse set conjecture*, volume 236 of *Memoirs of the American Mathematical Society*. American Mathematical Society, 2014.

[19] G. Sargsyan and N. Trang. *The largest Suslin axiom.* 2016.

[20] Grigor Sargsyan and Nam Trang. The exact consistency strength of generic absoluteness for universally Baire sets. 2019. Available at math.rutgers.edu/∼gs481/.

[21] Grigor Sargsyan and Nam Trang. Sealing from iterability. *Transactions of the American Mathematical Society, Series B*, 8(7):229–248, 2021.

[22] Jack Silver. Every analytic set is ramsey. *The Journal of Symbolic Logic*, 35(1):60–64, 1970.

[23] J.R. Steel and Nam Trang. Characterization of $\square$ in hod of AD^+ models. *In preparation*, 2022.

[24] Jok M Tang and Yousef Saad. A probing method for computing the diagonal of a matrix inverse. *Numerical Linear Algebra with Applications*, 19(3):485–501, 2012.

[25] N. Trang. PFA and guessing models. *Israel Journal of Mathematics*, pages 607–667, 2016.

[26] Nam Trang. HOD in natural models of AD^+. *Annals of Pure and Applied Logic*, 165(10):1533–1556, 2014.

[27] W. H. Woodin. *The axiom of determinacy, forcing axioms, and the nonstationary ideal*, volume 1 of *de Gruyter Series in Logic and its Applications*. Walter de Gruyter & Co., Berlin, 1999.