Abhishek Divekar's Step-by-Step guide to emailing professors for MS Thesis opportunities

In the following doc, I will share my thoughts on efficiently finding professors for research opportunities. I'll focus on ML professors (because that's what I know well).
I've tried to put this together in a step-by-step fashion, starting from the very basics.

Step 1. Timeline:

1. **Time commitment for finding advisors**: This whole process takes maybe 2-3 dedicated weeks of effort (this can be spread out if you start earlier). You need to shortlist professors whose interests match yours, learn what they do, then send an email expressing your ABILITY to do research in their sub-field (availability and excitement are nice but secondary things to express).
2. **Registration timelines**: before the 12th day of class, you must have (i) spoken to a professor and got them to agree to supervise (ii) email the grad coordinator, who will send an online form, which you and prof both submit to register for the Thesis. These two items must be completed before the 12th day of class each semester (same as the deadline for adding other courses). In Spring 2023, this was 25th Jan.
3. **Which semester?**: IMO you should start emailing professors 1 year before you want to graduate. E.g. I wanted to complete my final course in Fall 2023, so I spent Dec 2022 emailing professors, asking if they could take me on for Spring+Summer or Summer+Fall.

Why start 1 year in advance? Because professors often plan their research commitments many months ahead. These are private plans, so you can only guess when they will be busy. Suggesting Spring+Summer or Summer+Fall decreases your chances of getting a rejection because a professor is "too busy" (I will talk about other tips soon).
Additionally, even if you don't make the 12th-day-of-class deadline, a professor might be open to working together "unofficially" for the second half of a semester, and then advise your Thesis "officially" the following semester.

Step 2. Shortlisting advisors (5-10 hours):

An often-overlooked part of the process. **You can minimize your time spent reading research and drafting fruitless emails, if you aggressively shortlist advisors**. I personally did a lot of research on each potential advisor and sent only 2 highly-targeted emails, both of whom responded with options to chat (which turned into options to work on a research project).

1. **Very first thing**: see who you are allowed to advise your Thesis. Visit the GSC members list: (i) select "Computer Science" (ii) look at "Can be sole supervisor" = Yes. Surprisingly, you will see many faculty who are not in the CS department, but in affiliated departments like Statistics, iSchool, ECE, Linguistics, etc. You can also reach out to them as advisors! (Note: the "End Date" is the date when they are removed from the CS GSC list, but I assume they can renew it without issue).
    - If you are interested in ML research, UT MLL (ml.utexas.edu) lists anyone working on anything even distantly ML-related, so take an intersection of GSC and MLL as your starting point.
2. Next, go through this list, and ignore anyone who is an Emeritus professor (i.e. retired) or Adjunct professor (i.e. only teaches, no research).
3. **Start a Google doc** to capture various aspects of each professor as a potential advisor. Here is mine.
4. Visit each professor's webpage, and figure out, **at a very high level**, what their **research field is**.
    a. Do they work in ML, a different CS discipline, or another department altogether? Luckily, UT MLL (ml.utexas.edu) lists anyone working on anything even distantly ML-related, so this is simple. Side Note: ECE and iSchool departments are most similar to CS when it comes to ML...Statistics, Linguistics etc approach ML from a very different perspective than CS.
    b. If the prof works in ML, figure out which field: NLP, Computer Vision, AI Safety, Speech processing, Robotics, ML Theory, Reinforcement Learning, HCI? It is unlikely that they seriously work in more than one field: ML is so popular that each field has become a huge research space and profs must specialize. However, academics often dabble, and might have papers in multiple fields; see the field where the majority of their papers are published in the last 2 years.

    c. Look which conferences they tend to publish in; this is a strong indicator of the field they work in. If it's mostly ACL/EMNLP/COLING/TACL/etc, they work in NLP. If it's CVPR/ECCV/ICCV/WACV/etc, they work in Computer Vision. However, some prestigious conferences have become catch-all and are not good indicators of the field, e.g. NeurIPS / ICLR / ICML / KDD / WebConf / etc. If they publish mainly in journals, they don't work in CS (as conferences are the main research venues in CS).
5. The previous step should help you shortlist/remove profs based on relevant research fields.
    a. Personally, I was (i) keenly interested in NLP, (ii) could possibly make a case to work in Speech/CV/AI Safety/HCI, and (iii) had no interest or experience in Robotics/ML Theory/RL/Statistics research. I used this to narrow down a **"high-recall" list of 10-15 professors**, such that everyone worked in a field I was at least mildly excited about.
6. I then visited each professor's webpage and filled in the following columns in my Google sheet: "Name", "Email", "Tenure (Asst/Assoc/Full)", "Tenure at UT (years)", "Lab URL & focus", "Lab makeup".
    a. **Email**: look at their CV, not website (the CV will have a priority email address).
    b. **Tenure at UT**: look at their CV for this. If they have transferred from another university, note down both how long they have been a professor, and how long they have been at UT.
    c. **Tenure (Asst/Assoc/Full)**: Assistant professors are up for Tenure in their 6th year. Getting tenure is a super stressful process; an Assistant prof must lead research projects which establish a new paradigm in their sub-field. When a professor gets tenure, they are essentially "established". They become somewhat famous in their research sub-field, so their lab-size and commitments tend to grow, and even their PhD students find it difficult to get 1:1 time. Assistant or newly-Associate profs have fewer time-commitments and will probably be better advisors. However, a famous and established professor can write a recommendation letter which carries a lot of weight in PhD applications. So it's somewhat a tradeoff.
    d. **Lab URL & focus**: each professor's team of students and postdocs is called their "lab". The professor can call themselves the "Director", "Lead", "Principal

investigator (PI)", "Grand Emperor", whatever. Sometimes they give their labs fun names (e.g. TAUR lab, SALT lab, TRISHUL, etc).
More senior professors are directors of large, collaborative labs like UT MLL (ml.utexas.edu), and UT NLP (www.nlp.utexas.edu). Don't be fooled! These large labs are basically a Facebook group of like-minded people; figure out which PhD/MS students and postdocs each professor works with personally. This should be on a “personal lab” webpage, with around 7-12 current members.
Each professor's lab focuses on 3 or 4 hyper-specific topics within their broad research field. The lab's homepage will specifically call out what they focus on. E.g. Greg Durrett's TAUR lab works on "building models that have a greater ability to analyze, understand, and reason about the content of text". **This short summary is actually a crucial part of sending your email, since it informs you what the professor's general contribution to their field of research is**. Read this summary carefully a few times, and save it in your Google doc.

e. **Lab makeup**: visit each student in the professor's lab, and see how long they have been at UT. Assume they will graduate after their 5th year is complete. **Tabulate the statistics of the lab (how many 1st/2nd/3rd/4th/5th year PhD students, how many Masters/Undergrad students) to see how much bandwidth the professor will have for the semesters when you are proposing to be their student**. Note that every professor assigns bandwidth differently, so this is only an indication of their availability. But look out for opportunities; if they have many students graduating this year.
One of the reasons I got my thesis advisor was because they had bandwidth, and said so explicitly: "Hi Abhishek. Thanks for reaching out! I know you've reached out before and unfortunately I didn't have any bandwidth -- however, I have several students finishing this semester and so I might be able to find something."

7. If you're lucky, on their personal/lab webpage, a professor might mention if they are taking students, or how PhD applicants/MS students should apply for research opportunities. **Make sure to note down these instructions, and follow them exactly when reaching out**.

    a. If you see a professor is seeking PhD students, move them higher on your shortlist (this means they have bandwidth to advise PhDs, and are thus possibly open to advising Master's students too).
    b. Remove the professor from consideration if:
        i. They mention they are not taking students.
        ii. Their website has REALLY out of date information (they probably don't respond to emails).
        iii. They have greatly reduced their publishing in the last year or so (might be seeking another appointment). Use Google Scholar to determine this (webpage might be out of date).
    c. Note: some tenured professors take up **industry positions** (e.g. Inderjit Dhillon, Sujay Sanghavi and Matt Lease). These may be full-time (e.g. VP position at Amazon/Google/Microsoft/Meta) or part-time (e.g. Amazon Scholars program). In this case, it is ambiguous whether they are taking students, so look for other signs (like whether they had new PhD students who started last year), or check their Twitter.
8. On their webpage, hunt for the prof's Twitter profile, and read every Tweet in the last 2 years. I am dead serious. **AN ACTIVE TWITTER PROFILE IS BY FAR THE #1 RESOURCE FOR**: (i) what a professor is working on recently; (ii) their personality; (iii) are they looking for new students; (iv) which of their students are graduating.
    a. If you don't have a Twitter profile, ignore all your opinions about Elon Musk, and create a Twitter profile RIGHT NOW. The ML Twitter community is truly invaluable: majority of prominent ML academics use Twitter to share papers, conference acceptances, explanatory blogs, memes, their thoughts on different ML ideas, etc. It is more upto-date than anything else.
    b. For example, when Dr. Scott Neikum (RL course) decided to move from UT Austin to UMass Amherst, he announced it on Twitter (yes, Twitter!) months before updating his "official" website. C'est la vie.

By this point you should have looked at many websites and got a general idea of which profs work on which topics, without actually reading many research papers. The next step is documenting your own skills, which will help you further prioritize.

Step 3. CV and personal homepage:

You need to have a CV link, which you paste in your email to professors (never send your CV or Transcripts as an attachment! It seems spammy).
I'm not explaining the differences between a CV and a Resume, ask Google or ChatGPT.
Create a CV.
Do consider creating a personal website, I just slapped something together using AcademicPages + Github pages (here it is: https://adivekar-utexas.github.io/cv/).
Don't spend more than 2 days on both CV+homepage because to be brutally honest, no one cares how pretty it is, and barely anyone reads it. I have personally used "optimizing my CV" as a way to procrastinate Step 4 and 5; don't do it.

After this step, you should be able to re-order your list of advisors based on the closest match to your own skillset/projects. In your email, you will need to sell yourself in a few lines, and I recommend highlighting parts of your past experience that will be most attractive to the prof. E.g. I have done UT course projects on CV, NLP and Speech, but my day-to-day work deals with NLP. A CV professor will not be interested if I mention my NLP work in an email.

Step 4. Reading research:

You should now have a list of professors, ordered by relevance, that you can work through from top-to-bottom. This is good, but now you must understand what the hell they work on, in order to create the body of your email.

1. First, visit the Google sheet and re-read what their lab's Focus.
    a. If you still don't understand it, see if they have any recent "Invited Talks". These are talks where they typically explain (to a technical but unfamiliar audience) what their lab has been working on recently.
    b. You might find it useful to browse their Twitter, sometimes they explain some of the papers in plain English.
    c. If you really still don't understand, look at the Introductions of some of the papers.
2. You will come to the realization that within their research field (e.g. NLP/CV/AI Safety/etc) each professor has been working on 3-4 hyper-specific "research tracks"

in the last few years. E.g. Greg Durrett has been working on "checking and improving the factual correctness of generative NLP models for abstractive summarization", among other things. **Typically, one/two of their PhD students is the first-author of all papers in a specific track. Make careful note of which PhD students work on which track (it will be useful later).**

3. **Read papers on only one research track which you feel aligns well with your own CV**, since you will need to comment on it in your email, and mention how you want to extend it. Exploring only one research track in depth per professor, will save you a lot of time. In your email, mentioning too many tracks seems unfocused.
    a. You don't need to read the papers in TOO much technical depth; just enough to understand the key contributions of the work, and where it fits in the research landscape. Basically, read the Introduction and some of the references (see tips below).
    b. Sometimes, you will not understand an important part of the paper you need to read e.g. you might not know some technical detail about the T5 model. In this case, it's worth spending a day actually looking at scientific blogs or scientific YouTube videos to understand it. Some resources I recommend: lilianweng.github.io, d2l.ai (best recent book on Deep Learning), Kevin Murphy's 2022 ML books.

Here's a few practical tricks for reading ML research papers briefly:

1. **"Introduction Is All You need"**. The Title is a buzzword-filled acronym, the Abstract is an indecipherable mess, but the Introduction actually explains, in plain English, everything you need: what are the missing pieces in current research, what this paper contributes, and maybe some results. Definitely read the Introduction, and possibly the "Related Work" section if you want to understand the background a bit better.
2. **Top 3 references**: A paper might have fifty references, but there are typically only 2-3 papers which the work immediately builds on top of. Identify these references, and see if they repeat across the papers you have to read. If they do, figure out if you need to read them. They are possibly popular approaches/models, so see if there is a scientific blog which explains the core idea in simple English (don't actually read the referenced paper unless you have to).

3. **Contextualized highlighting**: When reading a paper, I have found it helps to highlight text as follows: "Green"=background knowledge (previous work, etc); "Yellow"=current paper (details of what the authors did in this paper); "Red"=key insights (something unintuitive the authors discovered, or the performance improvements, or the novel architecture proposed).
4. **Paper manager**: You will read many papers in this process and many more during your Thesis. It's useful to invest in a good paper-management software. I have personally tried 5-6 and found ReadCube Papers to be the absolute best cross-platform solution (UT email also gets a discount so that it is only $3/month). Other good ones are Zotero, Mendeley, Paperpile, Sciwheel. This is 100% worth paying for, rather than adopting a half-baked open-source alternative.

Step 5. Sending Emails:

Let's get this out of the way: cold-emailing sucks because you might spend 3-4 days reading papers, looking up numerous terms, actually get a bright idea of how to extend someone's research, spend hours crafting an email with your thoughts, nervously hit send, only to potentially get ghosted or a simple "sorry I'm busy" reject. I hated this overall experience, so I did several things to maximize my chances:

1. Sent it from my **@utexas.edu account**, to minimize chances of spam-filtering.
2. **Eye-catching subject line**: I tried "Master's Thesis for Amazon Applied Scientist" and "Thesis supervision for UT CS student"...I felt these would stand out from the other 100 emails they get from students wanting to work on research.
3. The first paragraph should be crisp and state your motivation, availability, and why you are sending this email. This can be common across all your emails. Read this blog on etiquette for emailing professors.
    a. This was my first paragraph: "Hi Dr. X, Hope you have had happy and healthy holidays. My name is Abhishek. I am a Master's student at UTCS; I also work full-time as an Applied Scientist at Amazon. I would like to enquire if you might consider supervising my Master's Thesis starting Spring 2023 or Summer 2023. My motivation is to produce a paper to be submitted to a top-tier NLP venue in 2023 (EMNLP, COLING, TACL, etc). It would be great if

you are open to discuss this in the first few weeks of January. My CV can be found here: https://adivekar-utexas.github.io/cv/".

4. **A note on "Motivation":** think hard about ***why*** you are doing the Thesis, as you will need to mention this in the 1st/2nd paragraph.
    a. **Good reasons**: want to do a PhD, or want to publish papers and move into a ML research job in industry.
    b. **Bad reason**: "ML research looks exciting and I'm interested in contributing" -> this indicates to a prof that once you realize research is hard, boring, and gut-wrenchingly ambiguous work, you might drop it altogether (having wasted everyone's valuable time).
    c. Don't try to lie about your motivation (e.g. say you want to do a PhD when you have not given it serious thought). This will come across as either fake or naive.
    d. Instead, reflect on why you feel this desire to do a Thesis. If it's FOMO, I can tell you from experience that an ML Engineering job is more exciting, easy, and financially rewarding than ML research. However, if you have a serious career reason to do a Thesis, express this in your email explicitly, e.g. "I hope to apply to PhD programs by the end of the year". People like people who are crystal-clear on why they are doing something.
5. **A note on email tone**: be polite and to-the-point. A useful statistic to know is that out of PhD holders, less than 10% actually manage to become professors. Even junior professors are rockstars in their sub-field. Many students want to work with them and send them emails expressing this continuously. That said, Professors in general are polite and nice people, just extremely busy. If you are rude or ask dumb questions (e.g. those already answered on their webpage), you will likely not get a response, or a generic "form email" response. See more in this Twitter thread. Your email does not need to take a reverential tone or sound "in awe" of the professor; be polite and to-the-point.
6. **2nd paragraph, opening**: I stated (i) where I was in the program and current GPA (ii) a relevant UT course I took, and my grade (when I emailed Greg Durrett, I said "I took your NLP course and got an A". I left this off when I emailed another NLP professor; during our meeting, she asked if I had taken Greg's NLP course).

    a. This was what came out: "About myself: I started my Master's in Fall 2020, while working. I have completed 6 of 10 courses with a GPA of 4.0 (including Greg Durrett's NLP course, where my grade was A). I wish to complete a Thesis before graduating (expected Jan 2024)."
7. **Remaining 2nd paragraph & 3rd paragraph**: this was more open-ended. It's best to bring up their research in a way that sounds organic. I used the following approach:
    a. I briefly described one of my projects at UT / full-time job, and how this led me to discover their research on XX.
    b. Next, I expressed how I loved the work by ***one particular student*** (see Step 4.) and expressed some of my own ideas on this, framed as research questions (RQ).
    c. Overall, it came out something like this: "At Amazon, my work focuses on XXX. In this context, I really loved the line of work by <PhD student> [4, 5] evaluating YYY in the presence of ZZZ. The learning that QQQ was quite surprising. This has applications in my work, because <reason>. A few follow-up research questions occur to me: (a) <RQ1> (b) <RQ2> (c) <RQ3>."
8. **Short concluding paragraph**: "Please let me know if you have space in your lab, as this line of research is really interesting to me. It would be great if we could connect in January to discuss further. Thank you for your time."
9. **References**: have you ever seen an email with a References section? Well, my emails did have this, and I thought it was a great idea for two reasons: (i) it shows you are familiar with writing in a research style (ii) it flows much better than copy-pasting the paper title inline in the paragraph. I added References when talking about my own work, when talking about their papers, and when talking about research questions.
10. **Editing**: my initial drafts were pretty long. I needed to cut, but did not know how much. So, I sent the draft email to myself and looked at it on my phone (I figured a busy professor might read mail on their phone). When I looked at it this way, it became clear that both my Subject and Body were too long. I made the Subject more snappy, and kept cutting the body-text until it fit roughly within 1-1.5 phone-screens.

Step 6. Follow-up Emails:

I personally hate bothering people with follow-ups, but a friend at UT convinced me it's absolutely required because professors are super busy and get literally hundreds of emails per day.

I decided I would do this in a structured way:

1. I sent the email using Gmail's Scheduled sent feature, so that it would appear in their inbox at roughly 9am on a weekday.
2. I copied the email into my Google doc, along with the send-date.
3. I immediately wrote out the follow-up email. It was really short: "Hi X, I was wondering if you had a moment to consider my request for Master's Thesis supervision (as per the email below). I'm very excited to dig into <research-topic-you-emailed-about>, but more than open to suggestions for other research topics. The research coming out of your lab is super interesting (particularly <Phd-student-1> and <Phd-student-2>'s work). Thank you for your time!"
4. I set up a Google reminder to send the follow-up email 5 days later (I was too nervous to use Gmail Scheduled Send, since the prof might actually respond and the follow-up email would look weird).
5. After another 5 days, if I did not get a response, ***I would write and send an email to the PhD student I had mentioned in the main email***.
    a. I would basically introduce myself briefly, copy-paste the Research Questions about their work, and include a small request at the bottom to ask the professor to check my email.
    b. The request was something like this: "Finally, a small request from me: I am completing my Masters in Jan 2024, and wanted to do a Thesis before graduating. I really enjoyed your work, and reached out to Dr. X last week, to see if I could work with her on a similar problem. I have not received a response (even after sending a reminder). If you could quickly ask Dr. X if she has seen my email, I would really appreciate it. I sent the mail from
    "

If after all these steps, I did not get a response, I just abandoned hope and moved on. One option is to reach out to the professor on Twitter, but I did not want to be that desperate.

Conclusion:

That's what I did! 😄I know it seems like sending these emails is some long, difficult process. But really, you just need to give some thought into targeting the **right** professors. I sent only 2 emails, but both professors met with me and could have been potential advisors, so maybe this sort of laser-sharp targeting was useful.