

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده علوم ریاضی

رشته ریاضی کاربردی، گرایش کنترل و بهینه سازی

رساله دکتری

حل رده ای از مسائل برنامه ریزی نیمه معین با استفاده از شبکه های عصبی

نگارنده: آسیه نیک سرشت

استاد راهنما

دکتر علیرضا ناظمی

بهمن ۱۳۹۷

تقدیم بہ خانوادہ عزیزم
کہ آرامشم را
و بہ استاد کرامتقدر جناب آقای دکتر ناظمی
کہ رسیدن بہ اہداف علمی ام را
مدیونشان ہستم

سپاس‌گزاری...^پ

خدایا! شکر که تو را حدی برایت متصور نیست. سپاس تو را که بی نهایت مهربانی. خدایا! حمد و سپاس از آن توست که محیطی بر تمام عالم و تو تکیه گاهی بر تمام موجودات. معبودا! تو آن قدر گویایی که من از وصف تو ناتوانم. بند بند وجودم از شعاع نور توست که گرم و روشن می شود.

اکنون که با عنایت خداوند متعال این دوره‌ی علمی به پایان رسیده است وظیفه خود را می‌دانم که تشکر قلبی خود را نسبت به کسانی که به طریقی در انجام این پژوهش مرا یاری نموده‌اند ابراز دارم. در آغاز وظیفه‌ی خود می‌دانم از زحمات بی‌دریغ استاد راهنمای خود، جناب آقای دکتر علیرضا ناظمی صمیمانه تشکر و قدردانی کنم که از راهنمایی‌های ارزنده ایشان در راستای پیشبرد پژوهش حاصل فراوان بردم و همواره شاگرد مکتب علم و انسانیت و منش والای ایشان هستم.

همچنین لازم می‌دانم از اساتید فرهیخته جناب آقای دکتر سهراب عفتی، جناب آقای دکتر حجت احسنی طهرانی و جناب آقای دکتر مهرداد غزنوی که داوری این پایان‌نامه را به عهده گرفتند با تمام وجود تشکر و قدردانی نمایم. در پایان از همه دوستان و عزیزانی که با اندیشه و قلبشان در این راه یاری‌ام نمودند، خالصانه تشکر و قدردانی می‌نمایم.

آسیه نیک سرشت

بهمن ۱۳۹۷

تعهد نامه

اینجانب آسیه نیک سرشت دانشجوی دکتری رشته ریاضی کاربردی علوم ریاضی دانشگاه شاهرود، نویسنده پایان نامه با عنوان حل رده ای از مسائل برنامه ریزی نیمه معین با استفاده از شبکه های عصبی، تحت راهنمایی علیرضا ناظمی متعهد می شوم:

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش های دیگر پژوهش گران، به مرجع مورد استفاده استناد شده است.
- مطالب این پایان نامه، تا کنون توسط خود، یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارایه نشده است.
- حقوق معنوی این اثر، به دانشگاه صنعتی شاهرود تعلق دارد، و مقالات مستخرج با نام “ دانشگاه صنعتی شاهرود “ یا “ Shahrood University of Technology “ به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آوردن نتایج اصلی پایان نامه تاثیرگذار بوده اند، در مقالات مستخرج از پایان نامه رعایت می گردد.
- در تمام مراحل انجام این پایان نامه، در مواردی که از موجود زنده (یا بافت های آنها) استفاده شده است، ضوابط و اصول اخلاقی رعایت شده است.
- در تمام مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته (یا استفاده شده است)، اصل رازداری و اصول اخلاق انسانی رعایت شده است.

آسیه نیک سرشت

بهمن ۱۳۹۷

مالکیت نتایج و حق نشر

- تمام حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه های رایانه ای، نرم افزارها و تجهیزات ساخته شده) متعلق به دانشگاه صنعتی شاهرود می باشد. این مطلب باید به نحو مقتضی، در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در این پایان نامه بدون ذکر منبع مجاز نمی باشد.

چکیده

مسائل بهینه‌سازی نیمه‌معین در دهه‌های اخیر بسیار مورد توجه قرار گرفته است و شامل حوزه بسیار گسترده‌ای از مسائل بهینه‌سازی محدب است. بسیاری از مسائل مهم در نظریه سیستم و کنترل، بهینه‌سازی ساختاری، آمار و سایر زمینه‌ها می‌توانند به صورت یک مساله برنامه‌ریزی نیمه معین طرح شوند. تاکنون روش‌های مختلفی برای حل مساله نیمه معین ارائه شده‌اند. اما استفاده از شبکه‌های عصبی برای حل این مساله به جز در تعداد انگشت‌شماری گزارش نشده است.

در این رساله دو مدل شبکه عصبی برای حل مسئله برنامه‌ریزی نیمه معین ارائه خواهد شد. نشان خواهیم داد که نقطه تعادل شبکه عصبی پیشنهاد شده و جواب بهینه مساله نیمه معین خطی معادل هستند. همچنین پایداری و همگرایی مدل‌های شبکه عصبی ارائه شده اثبات خواهد شد. همچنین برای اولین بار یک مدل شبکه عصبی برای حل مساله نیمه معین غیرخطی ارائه می‌دهیم. پایداری و همگرایی سراسری به جواب بهینه مساله اصلی اثبات می‌شود. با ارائه مثال‌هایی نتایج نظری را تایید می‌کنیم.

کلمات کلیدی: شبکه‌های عصبی، برنامه‌ریزی نیمه معین، برنامه‌ریزی محدب، همگرایی

سراسری

لیست مقالات مستخرج از پایان نامه -

۱. مقاله اول:

Asiye Nikseresht , Alireza Nazemi , A novel neural network model for solving a class of nonlinear semidefinite programming problems

Journal of Computational and Applied Mathematics 338 (2018) 69–79

۲. مقاله دوم:

Asiye Nikseresht , Alireza Nazemi, A novel neural network for solving semidefinite programming problems with some applications

Journal of Computational and Applied Mathematics 350 (2019) 309–323.

فهرست مطالب

ک فهرست تصاویر

س فهرست جداول

۱ مقدمات و مفاهیم اساسی

- ۱.۱ مقدمه‌ای بر شبکه‌های عصبی
- ۱.۱.۱ شبکه‌های عصبی بیولوژیکی
- ۲.۱.۱ شبکه‌های عصبی مصنوعی
- ۳.۱.۱ تاریخچه تکامل شبکه عصبی
- ۴.۱.۱ تاریخچه حل مسائل بهینه‌سازی با استفاده از شبکه‌های عصبی
- ۲.۱ مقدمه‌ای بر بهینه‌سازی غیرخطی
- ۳.۱ سیستم‌های دینامیکی
- ۴.۱ نامساوی‌های وردشی

۲ مقدمه ای بر مسائل بهینه سازی نیمه معین

- ۱.۲ مقدمه
- ۲.۲ تعاریف و نمادها
- ۳.۲ برنامه ریزی نیمه معین
- ۴.۲ مثال‌ها
- ۱.۴.۲ برنامه‌ریزی درجه دوم به صورت درجه دوم مقید
- ۲.۴.۲ بزرگترین مقدار ویژه و مینیمم‌سازی نرم ماتریسی
- ۳.۴.۲ تقریب لگاریتمی چبیشف
- ۴.۴.۲ بهینه‌سازی ساختاری
- ۵.۴.۲ کاربرد در آمار
- ۵.۲ دوگانی در برنامه‌ریزی نیمه معین

۳۷	یک مدل شبکه عصبی برای حل مسئله بهینه سازی نیمه معین خطی	۳
۳۷ مقدمه	۱.۳
۳۹ مدل سیستم دینامیکی	۲.۳
۴۰ تحلیل پایداری و همگرایی	۳.۳
۴۴ مثال های عددی	۴.۳
۵۳	یک مدل شبکه عصبی گرادیانی برای حل مسئله نیمه معین خطی	۴
۵۳ مقدمه	۱.۴
۵۴ یک مدل شبکه عصبی گرادیانی	۲.۴
۵۷ تحلیل پایداری و همگرایی	۳.۴
۵۹ مثال های عددی	۴.۴
۶۵	یک مدل شبکه عصبی برای حل مساله نیمه معین غیر خطی	۵
۶۵ مقدمه	۱.۵
۶۶ مدل شبکه عصبی	۲.۵
۶۸ تحلیل پایداری و همگرایی	۳.۵
۷۳ مثال های عددی	۴.۵
۷۷ نتایج و پیشنهادات	۵.۵
۷۹	مراجع	

فهرست تصاویر

۳	۱.۱	ساختار یک نرون.
۳	۲.۱	مدل ریاضی یک نرون.
۴	۳.۱	ساختار یک نرون با لایه ورودی، لایه پنهان و لایه خروجی.
۶	۴.۱	شبکه‌های عصبی دو لایه بازگشتی.
۱۳	۵.۱	پایداری لیپانوف.
۱۴	۶.۱	پایداری مجانبی.
۱۹	۷.۱	تعبیر هندسی مسأله نامساوی وردشی.
		۱.۳	رفتار همگرایی $(x_1(t), x_2(t), x_3(t), x_4(t))^T$ شبکه عصبی (۴.۳) و (۵.۳)
۴۵		در مثال ۱.۳.
		۲.۳	رفتار همگرایی $(x_1(t), x_2(t))^T$ از شبکه عصبی (۴.۳) و (۵.۳) با ۴ نقطه
۴۶		آغازین تصادفی در مثال ۲.۳.
		۳.۳	نمودار فاز $(x_1(t), x_2(t))^T$ از شبکه عصبی (۴.۳) و (۵.۳) با ۱۸ نقطه آغازین
۴۶		در مثال ۲.۳.
		۴.۳	رفتار همگرایی $(\lambda(t), x(t)^T)^T$ از شبکه عصبی (۴.۳) و (۵.۳) با ۵ نقطه
۴۸		آغازین تصادفی در مثال ۳.۳.
		۵.۳	رفتار همگرایی $\ (\lambda(t), x(t)) - (\lambda^*, x^*)\ ^2$ با $(\lambda_0, x_0^T)^T = (2, -1, 1, -1, 0, 1)^T$
۴۸		در مثال ۳.۳.
		۶.۳	رفتار همگرایی $(\lambda, \mu, x^T)^T$ شبکه عصبی (۴.۳) و (۵.۳) با ۵ نقطه آغازین
۴۹		تصادفی در مثال ۴.۳.
۴۹	۷.۳	رفتار همگرایی تابع هدف با ۱۰ نقطه آغازین تصادفی در مثال ۴.۳.
		۸.۳	رفتار همگرایی $\ x(t) - x^*\ ^2$ با $(\lambda_0, \mu_0, x_0^T)^T = (1, -1, 1, -1, 0, 2, 1)^T$ در
۵۰		مثال ۴.۳.
		۹.۳	رفتار همگرایی $(\nu(t), x(t)^T)^T$ در شبکه عصبی (۴.۳) و (۵.۳) با ۵ نقطه
۵۱		آغازین تصادفی در مثال ۵.۳.

	۱۰.۳ رفتار همگرایی $\ \nu(t), x(t)^T\ ^2 - (\nu^*, x^{*T})^T$ با $(\nu_0, x_0^T)^T = (2, 1, 2, -2, 1, 1)^T$ در	
۵۱ مثال ۵.۳	
	۱۱.۳ رفتار همگرایی $(x_1(t), x_2(t), x_3(t))^T$ از شبکه عصبی (۴.۳) و (۵.۳) با ۵	
۵۲ نقطه آغازین تصادفی در مثال ۶.۳	
۵۲ رفتار همگرایی $Z(t)$ از شبکه عصبی (۴.۳) و (۵.۳) در مثال ۶.۳	
	۱۳.۳ رفتار همگرایی $\ x(t) - x^*\ ^2$ مبتنی بر شبکه عصبی (۴.۳) و (۵.۳) با ۲۰	
۵۲ نقطه آغازین تصادفی و $\rho = 10$ در مثال ۶.۳	
	۱.۴ رفتار همگرایی $(x_1(t), x_2(t))^T$ از شبکه عصبی (۱۳.۴) و (۱۴.۴) با ۱۰ نقطه	
۶۰ آغازین تصادفی در مثال ۱.۴	
	۲.۴ رفتار همگرایی $(\lambda(t), x(t)^T)^T$ از شبکه عصبی (۱۳.۴) و (۱۴.۴) با ۲۰ نقطه	
۶۰ آغازین تصادفی در مثال ۲.۴	
	۳.۴ رفتار همگرایی $\ (\lambda(t), x(t)) - (\lambda^*, x^*)\ ^2$ با $(\lambda_0, x_0^T)^T = (2, -1, 1, -1, 0, 1)^T$	
۶۱ در مثال ۲.۴	
	۴.۴ رفتار همگرایی $(\lambda, \mu, x^T)^T$ شبکه عصبی (۱۳.۴) و (۱۴.۴) با ۲۰ نقطه	
۶۲ آغازین تصادفی در مثال ۳.۴	
۶۲ رفتار همگرایی تابع هدف با ۲۰ نقطه آغازین تصادفی در مثال ۳.۴	
	۶.۴ رفتار همگرایی $\ x(t) - x^*\ ^2$ با $(\lambda_0, \mu_0, x_0^T)^T = (1, -1, 1, -1, 0, 2, 1)^T$ در	
۶۳ مثال ۳.۴	
	۷.۴ رفتار همگرایی $(\nu(t), x(t)^T)^T$ در شبکه عصبی (۱۳.۴) و (۱۴.۴) با ۵ نقطه	
۶۳ آغازین تصادفی در مثال ۴.۴	
	۸.۴ رفتار همگرایی $\ \nu(t), x(t)^T\ ^2 - (\nu^*, x^{*T})^T$ با $(\nu_0, x_0^T)^T = (2, 1, 2, -2, 1, 1)^T$ در	
۶۴ مثال ۴.۴	
	۱.۵ رفتار همگرایی $(x_1(t), x_2(t), x_3(t), x_4(t))^T$ شبکه عصبی (۱۱.۵) و (۱۲.۵)	
۷۴ با ۱۰ نقطه که به طور تصادفی در مثال ۱.۵ انتخاب شدند	
	۲.۵ نمودار فاز $(x_1(t), x_2(t))^T$ شبکه عصبی (۱۱.۵) و (۱۲.۵) با ۱۵ نقطه آغازین	
۷۴ در مثال ۱.۵	
	۳.۵ نمودار فاز $(x_3(t), x_4(t))^T$ شبکه عصبی (۱۱.۵) و (۱۲.۵) با ۱۵ نقطه آغازین	
۷۵ تصادفی در مثال ۱.۵	
۷۵ رفتار همگرایی $\ x(t) - x^*\ ^2$ با $x_0 = (1, -1, 2, 1)^T$ در مثال ۱.۵	
۷۵ رفتار همگرایی $\ x(t) - x^*\ ^2$ با ۲۰ نقطه آغازین تصادفی و $\rho = 1$ در مثال ۱.۵	
	۶.۵ رفتار همگرایی $(x_1(t), x_2(t))^T$ مبتنی بر شبکه عصبی (۱۱.۵) و (۱۲.۵) با	
۷۶ ۵۰ نقطه آغازین تصادفی در مثال ۲.۵	

۷.۵ رفتار همگرایی $\|x(t) - x^*\|^2$ مبتنی بر شبکه عصبی (۱۱.۵) و (۱۲.۵) با ۲۰
نقطه آغازین تصادفی و $\rho = ۵$ در مثال ۲.۵ ۷۶

فهرست جداول

۵	۱.۱ برخی توابع فعال سازی
---	-------	--------------------------

فصل ۱

مقدمات و مفاهیم اساسی

۱.۱ مقدمه‌ای بر شبکه‌های عصبی

۱.۱.۱ شبکه‌های عصبی بیولوژیکی

واحد اصلی دستگاه عصبی، سلولی خاص به نام نرون^۱ می‌باشد و بدون شک نحوه کار مغز و راز شعور آدمی در آن نهفته است. نرون‌ها با وجود تفاوت‌های زیاد از نظر اندازه و شکل ظاهری، مشخصه‌های مشترکی دارند. همان‌طور که در شکل ۱.۱ نشان داده شده است از تنه سلولی نرون تعدادی شاخک کوتاه خارج می‌شود که دندریت^۲ نام دارند. دندریت‌ها و تنه سلولی، سیگنال‌ها را از نرون‌های مجاور دریافت می‌کنند و از طریق یک لوله باریک به نام آکسون^۳ به نرون‌های دیگر منتقل می‌گردد.

آکسون در انتهای خود به تعدادی رشته جانبی باریک تقسیم می‌گردد که پایانه‌های آکسونی نام دارد و با دندریت‌های سایر نرون‌ها مرتبط است. ارتباط بین آکسون یک نرون و دندریت نرون دیگر را سیناپس^۴ می‌نامند. یک سیناپس مرکب از پایانه قبل از سیناپسی^۵، شکاف

^۱Neuron

^۲Dendrite

^۳Axon

^۴Synapse

^۵ Presynaptic Terminal

سیناپسی^۶ و پایانه بعد از سیناپسی^۷ می باشد. همه سیگنال های جمع بندی شده، در تنه نرون ترکیب می شوند و اگر وسعت سیگنال های ترکیب شده به آستانه نرون برسد، مرحله تحریک شدن^۸ فعال می شود و یک سیگنال خروجی تولید می شود. این سیگنال به صورت یک پالس منفرد یا بخشی از پالس ها در یک میزان خاص به موازات آکسون به پایانه های سیناپسی انتقال می یابد و موجب ترشح موادی به نام عصب-رسانه می شود. عصب-رسانه در داخل شکاف سیناپسی پخش می شود و نرون بعدی را تحریک می کند و به این ترتیب یک سیگنال از یک نرون به دیگری انتقال می یابد. تعداد بسیار زیادی آکسون از نرون های مختلف با دندریت های یک نرون به این صورت ارتباط برقرار می کنند.

نرون ها با توجه به عملکردشان به سه دسته تقسیم می شوند:

(۱) **نرون های حسی:** این نرون ها پیام ها را از گیرنده ها به دستگاه عصبی مرکزی منتقل می کنند. گیرنده سلول خاصی است که در اندام های حسی، عضلات، پوست و ... قرار دارد و تغییرات فیزیکی یا شیمیایی را در می یابد، این تغییرات را به سیگنال های عصبی تبدیل می کند.

(۲) **نرون های حرکتی:** این نرون ها پیام ها را از مغز یا طناب نخاعی به اندام های عمل کننده می رسانند.

(۳) **نرون های رابط:** این نرون ها پیام ها را از نرون های حسی می گیرند و به یک نرون رابط دیگر انتقال می دهند و یا پیام ها را به یک نرون حرکتی می رسانند. نرون های رابط فقط در مغز، چشم و طناب نخاعی یافت می شوند.

حال منظور از یادگیری عبارت است از تغییری نسبتاً ساختاری در سیناپس (بعد از تحریک)، که این تغییر ساختاری باعث افزایش کارایی سیناپس می شود.

۲.۱.۱ شبکه های عصبی مصنوعی

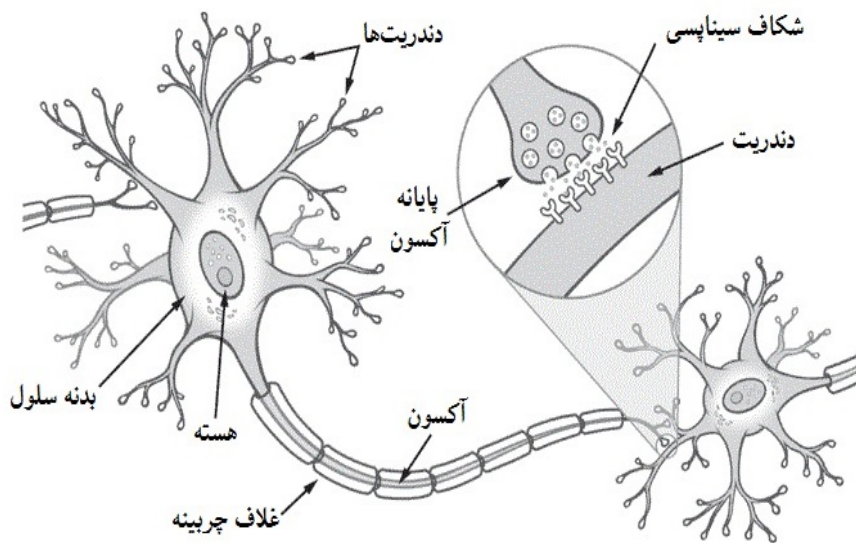
بعد از اشاره ای گذرا بر شبکه عصبی بیولوژیکی و نشان دادن گوشه ای بسیار کوچک از این جهان مرموز و فوق العاده پیچیده به موضوع اصلی یعنی شبکه های عصبی مصنوعی^۹ می پردازیم. شبکه های عصبی مصنوعی مدل های سخت افزاری یا نرم افزاری الهام گرفته از ساختار و رفتار نرون های بیولوژیکی و سیستم های عصبی انسان هستند که شامل عناصر پردازشی (به نام نرون ها) و ارتباطات میان آنها با ضرایب (وزن ها) اختصاص یافته به ارتباطات است، که ساختار شبکه عصبی را تشکیل می دهند. در این پایان نامه ما شبکه های عصبی مصنوعی را یک مدل

^۶ Synaptic Cleft

^۷ Postsynaptic Terminal

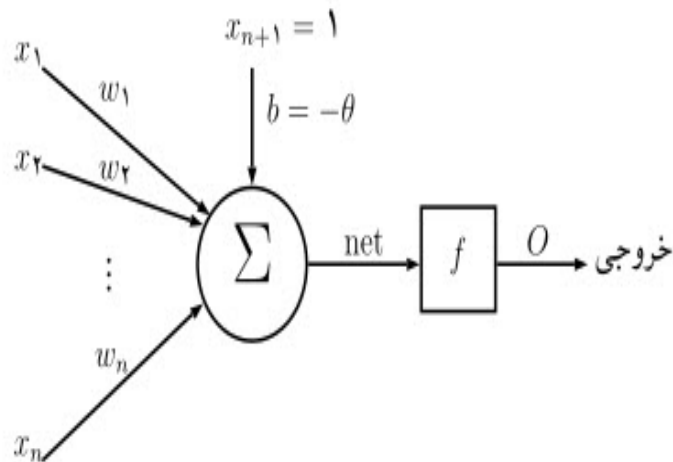
^۸ Firing

^۹ Artificial Neural Networks



شکل ۱.۱: ساختار یک نرون.

ریاضی فرض خواهیم کرد و به جای استفاده از عبارت شبکه‌های عصبی مصنوعی به طور خلاصه عبارت شبکه‌های عصبی را به کار خواهیم برد.

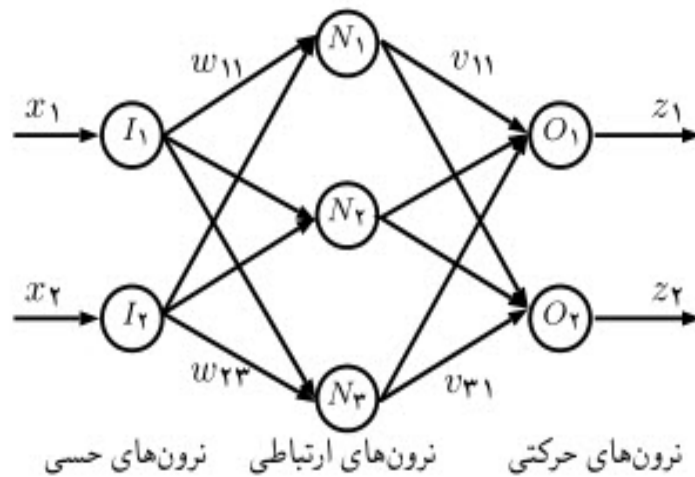


شکل ۲.۱: مدل ریاضی یک نرون.

مدل شبیه‌سازی شده یک نرون، در شکل ۲.۱ نمایش داده شده است. یک نرون در حالت کلی مجموعه‌ای از $n+1$ ورودی x_j ($j = 1, \dots, n+1$) دارد. این ورودی‌ها همان سیگنال‌های عصبی می‌باشند که از محیط یا از یک نرون دیگر فرستاده می‌شوند. w_j ($j = 1, \dots, n+1$) که روی آکسون‌ها واقع هستند نقش ساختار سیناپسی را ایفا می‌کنند. این‌ها هستند که باید تغییر کنند تا فرآیند یادگیری اتفاق بیفتد حال این که این تغییر بر چه قاعده‌ای استوار باشد بستگی به الگوریتم آموزش یا تعلیم دارد. وزن آخرین ورودی یعنی θ را با b نشان می‌دهند و آن را بایاس می‌نامند. سیگنال‌های عصبی x_1 و x_2 و ... و x_{n+1} به ترتیب در w_1, w_2, \dots

w_{n+1} ضرب شده و مجموع این حاصل ضرب‌ها که در شکل آن را با net نشان می‌دهیم، وارد نرون می‌شوند. در این مرحله نرون همانند یک تابع روی net عمل می‌کند و حاصل کار خود (برد تابع) را از طریق آکسون‌های طرف راست به نرون‌های دیگر می‌فرستد. این تابع را تابع فعال‌سازی^{۱۰} می‌نامند.

البته خاطر نشان می‌کنیم که اگر نرون به‌عنوان تابع فرض شود این تابع از \mathbb{R}^n به \mathbb{R} تعریف می‌گردد. یعنی سیگنال خروجی از نرون در تمام آکسون‌های طرف راست یکسان می‌باشد که روی شکل با O نمایش داده‌ایم. همان‌طور که قبلاً اشاره شد نرون‌ها می‌توانند حسی، ارتباطی و یا حرکتی باشند. اطلاعات از طریق نرون‌های حسی دریافت می‌شود و از طریق نرون‌های ارتباطی انتقال می‌یابد و بالاخره نتیجه کار شبکه عصبی و یا واکنش شبکه عصبی به اطلاعات دریافت شده از طریق نرون‌های حرکتی به خارج منتقل می‌گردد (شکل ۳.۱).



شکل ۳.۱: ساختار یک نرون با لایه ورودی، لایه پنهان و لایه خروجی.

نرون‌های حسی را لایه ورودی، نرون‌های حرکتی را لایه خروجی و نرون‌های ارتباطی را لایه پنهان نیز می‌گویند. لایه مخفی می‌تواند از چندین لایه تشکیل گردد البته در شکل ۳.۱ لایه مخفی تک لایه در نظر گرفته شده است.

مشخصه‌های اصلی یک شبکه عصبی که در طراحی آن باید مدنظر قرار گیرند عبارت‌اند از

معماری شبکه عصبی: نحوه اتصالات بین نرون‌ها، تعداد آن‌ها و تعداد لایه‌های تشکیل دهنده بخش نرون‌های ارتباطی را معماری شبکه عصبی می‌گوییم.

تابع‌های فعال‌سازی: این که چه تابعی روی نرون قرار بگیرد تا بر ورودی‌های نرون اثر کرده و خروجی نرون را تولید کند تابع فعال‌سازی آن نرون گویند. در جدول ۱.۱ تعدادی از مهم‌ترین توابع فعال‌سازی نشان داده شده است.

^{۱۰} Activation function

الگوریتم آموزش: روشی که وزن‌های روی آکسون‌ها بر اساس آن تغییر می‌کنند (w_{ij} ها و v_{jk} ها در شکل ۳.۱) تا خروجی شبکه عصبی به حالت مطلوب درآید را الگوریتم آموزش می‌گوییم.

جدول ۱.۱: برخی توابع فعال‌سازی

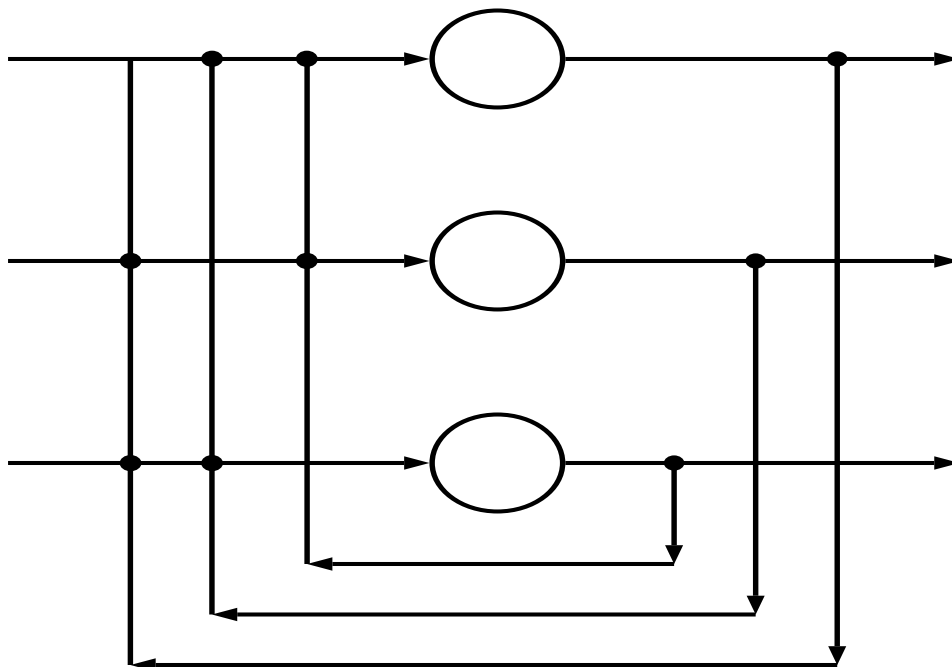
شکل تابع	تعریف تابع	نام	ردیف
	$f(x) = \begin{cases} 0 & x < 0, \\ 1 & x \geq 0. \end{cases}$	آستانه‌ای دو مقداره	۱
	$f(x) = \begin{cases} -1 & x < 0, \\ 1 & x \geq 0. \end{cases}$	آستانه‌ای دو مقداره متقارن	۲
	$f(x) = x$	خطی	۳
	$f(x) = \begin{cases} -1, & x < -1, \\ x, & -1 \leq x \leq 1, \\ 1, & x > 1. \end{cases}$	آستانه‌ای خطی متقارن	۴
	$f(x) = \begin{cases} 0, & x < 0, \\ x & 0 \leq x \leq 1, \\ 1, & x > 1. \end{cases}$	آستانه‌ای خطی	۵
	$f(x) = \frac{1}{1+e^{-x}}$	سیگموئید	۶

با توجه به مطالب قبل می‌توان برای شبکه‌های عصبی مصنوعی خصوصیات عمومی زیر را نسبت داد:

۱. پردازش اطلاعات در عناصر بسیار ساده به نام نرون (سلول عصبی) انجام می‌گیرد.
۲. سیگنال‌های عصبی بین سلول‌های عصبی از طریق اتصالات بین آن‌ها مبادله می‌شوند.
۳. به هر خط اتصال (آکسون) یک وزن نسبت داده می‌شود که در شبکه‌های عصبی معمولی این وزن‌ها در سیگنال‌های منتقل شده ضرب می‌گردند.
۴. هر نرون دارای یک تابع فعال‌سازی می‌باشد که این تابع در اکثر موارد غیرخطی است.

شبکه‌های عصبی بازگشتی

در شبکه‌های عصبی بازگشتی^{۱۱}، حداقل یک سیگنال برگشتی از یک نرون به همان نرون یا نرونهاي همان لایه و یا لایه قبل وجود دارد. چنین شبکه‌ای حافظه‌ای را نگه می‌دارد و حالت بعد نه تنها به سیگنال‌های ورودی بلکه به حالات قبل شبکه نیز وابسته است. این شبکه‌های عصبی معمولاً در مدل ریاضی به صورت معادلات دیفرانسیل مرتبه اول ظاهر می‌شوند.



شکل ۴.۱: شبکه‌های عصبی دو لایه بازگشتی.

^{۱۱}Feedback (Recurrent) Neural Network

۳.۱.۱ تاریخچه تکامل شبکه عصبی

در سال ۱۹۴۳ یک نرولوژیست به نام وارن مک کلاچ^{۱۲} به همراه یک متخصص آمار به نام والتر پیتز^{۱۳} اولین مدل ریاضی نرون را ارائه دادند [۶۲]. این نرون ساده یک عنصر محاسبه‌گر بود که عناصری به نام ورودی را در مقادیر ثابتی به نام وزن ضرب می‌کرد و آنها را از یک عملگر خطی عبور می‌داد و حاصل خروجی نرون را تشکیل می‌داد. در این مدل وزن‌ها مقادیر ثابتی بودند. در سال ۱۹۴۹ دونالد هب^{۱۴} [۳۳] کتابی درباره نحوه یادگیری در مغز انسان نوشت. بر این مبنا قانون یادگیری برای نرون را ارائه داد که به قانون هب معروف است و بعدها در تعلیم شبکه‌های عصبی از آن استفاده شد. در سال ۱۹۵۴ فارلی و کلارک^{۱۵} [۱] مدلی از نرون‌ها را به‌طور تصادفی به هم وصل کردند و قانون هب را پیاده‌سازی نمودند و نشان دادند که می‌توان دو الگوی ورودی را با این قانون تعلیم از هم تشخیص داد. با پیشرفت علم کامپیوتر اولین مدل از نرون عصبی مصنوعی توسط ناتانیال روچستر^{۱۶} محقق شرکت IBM در سال ۱۹۵۸ با استفاده از کامپیوتر شبیه‌سازی شد. روزنبلات^{۱۷} [۸۲، ۸۳] در سال ۱۹۵۸ نرون ساده مک کلاچ پیتز را اصلاح کرد و به آن قابلیت یادگیری و سازگاری را اضافه کرد و این نرون را پرسپترون نامید. قانون تعلیم پرسپترون^{۱۸} اولین قانون رسمی برای شبکه‌های عصبی است.

برنارد ویدرو^{۱۹} و مارسین هاف^{۲۰} و همکارانشان یک مدل سه سطحی از نرون‌ها را ابداع کردند که آدلاین^{۲۱} نام‌گذاری شد و مختصر شده حروف "عنصر خطی تطبیقی"^{۲۲} بود [۹۴، ۹۵، ۹۶، ۹۷] و برای آدلاین یک قانون تعلیم بر مبنای مشتق‌گیری بیان نمودند. این قانون تعلیم به $\alpha - LMS$ معروف است. از متصل کردن چند آدلاین به یکدیگر مادالاین^{۲۳} ساخته شد. ویدرو و هاف از شبکه عصبی مادالاین برای حذف اکو از خطوط تلفن استفاده نمودند.

سال ۱۹۶۹ آغاز یک دوره رکود برای شبکه عصبی بود. پاپرت و مینسکی^{۲۴} [۶۴] کتابی به نام پرسپترون نوشتند و نشان دادند که اگرچه گیت‌های OR و AND را می‌توان با استفاده از پرسپترون پیاده‌سازی کرد ولی گیت XOR قابل پیاده‌سازی نمی‌باشد و لذا نمی‌توان از

^{۱۲} Warren McCulloch

^{۱۳} Walter Pitts

^{۱۴} Donald Hebb

^{۱۵} Farley and Clark

^{۱۶} Nathaniel Rochester

^{۱۷} Frank Rosenblatt

^{۱۸} Perceptron

^{۱۹} Bernard Widrow

^{۲۰} Marcian Hoff

^{۲۱} Adeline

^{۲۲} Adaptive linear elements

^{۲۳} Madaline

^{۲۴} Papert and Minsky

شبکه عصبی برای پردازش اطلاعات استفاده کرد. الگوریتم پس انتشار^{۲۵} پال ورباس^{۲۶} [۹۳] در سال ۱۹۷۴ ارائه شد و بعدها توسط رامل هارت^{۲۷} [۸۴] به طور مستقل کشف گردید. این الگوریتم از زمان پیدایش به طور گسترده به عنوان یک الگوریتم آموزش در شبکه‌های عصبی پیش‌خورد^{۲۸} مورد استفاده قرار گرفته است. سال ۱۹۸۲ تا ۱۹۸۶ را می‌توان تولد دوباره شبکه عصبی دانست. در این سال‌ها دو اتفاق مهم در این زمینه رخ داد. اول ارائه شبکه عصبی بازگشتی و مفهوم تابع انرژی توسط جان هاپفیلد^{۲۹} [۳۶] بود که در سال ۱۹۸۲ منتشر شد. دومین اتفاق، حل مسأله فروشنده دوره‌گرد (TSP) توسط شبکه هاپفیلد بود [۳۷]. هاپفیلد نشان داد که از این شبکه بازگشتی می‌توان برای حل مسائل بهینه‌سازی استفاده کرد و افق جدیدی را در شبکه‌های عصبی مصنوعی گشود.

۴.۱.۱ تاریخچه حل مسائل بهینه‌سازی با استفاده از شبکه‌های عصبی

طیف گسترده‌ای از مسائل مهم در رشته‌های علوم پایه و مهندسی از جمله کنترل بهینه، طراحی ساختار بهینه، پردازش تصویر، تقریب توابع، تحلیل رگرسیون و... قابل تبدیل به مسائل بهینه‌سازی خطی و غیرخطی هستند. در دهه‌های گذشته الگوریتم‌های بهینه‌سازی عددی فراوانی برای حل مسائل بهینه‌سازی خطی و غیرخطی ارائه شده‌اند. از آن‌جا که زمان مورد نیاز برای حل مسائل بهینه‌سازی و به خصوص مسائل بهینه‌سازی غیرخطی بسیار وابسته به بعد و ساختار مسأله است بنابراین الگوریتم‌های عددی کارایی کمی از خود نشان می‌دهند. یک رهیافت امیدوار کننده برای حل مسائل بهینه‌سازی با ابعاد بالا استفاده از شبکه‌های عصبی مصنوعی می‌باشد.

در سال ۱۹۸۶ تانک و هاپفیلد [۸۶] اولین شبکه عصبی را برای حل مسائل برنامه‌ریزی خطی معرفی کردند و نشان دادند که وضعیت شبکه در هر تکرار به گونه‌ای تغییر می‌کند که تابع انرژی متناظر با آن به طور یکنواخت کاهش می‌یابد تا جایی که به نقطه مینیمم خود می‌رسد و این نقطه مینیمم متناظر با نقطه تعادل شبکه عصبی می‌باشد. آن‌ها این شبکه عصبی را توسط یک مدار الکتریکی پیاده‌سازی کردند. هم‌چنین از این شبکه برای حل مسأله فروشنده دوره‌گرد با ۳۰ شهر استفاده نمودند. این شبکه دارای نقص‌هایی بود به خصوص این که نقطه تعادل شبکه در شرایط کارش کان تا کر^{۳۰} صدق نمی‌کرد و لذا جواب مطلوبی از مسأله حاصل نمی‌شد. با این وجود کارهای هاپفیلد انگیزه بسیار خوبی را برای محققین به وجود آورد تا در این زمینه فعالیت کنند. کندی و چا^{۳۱} [۴۹] با افزودن یک پارامتر جریمه

^{۲۵} Back propagation

^{۲۶} Paul Werbos

^{۲۷} Rumelhart

^{۲۸} Feed forward

^{۲۹} John Hopfield

^{۳۰} Karush-Kuhn-Tucker (KKT)

^{۳۱} Kennedy and Chua

متناهی به شبکه هاپفیلد آن را توسعه دادند و از آن برای حل مسأله برنامه‌ریزی غیرخطی محدب استفاده نمودند. با پارامترهای جریمه این شبکه در یافتن نقطه بهینه دقیق ناتوان است به‌خصوص اگر پارامتر جریمه بزرگ باشد این شبکه به‌سختی عمل می‌کند.

برای جلوگیری از بکار بردن پارامتر جریمه یک شبکه عصبی سوئیچ-خازن^{۳۲} توسط رودریگز-وازکز^{۳۳} و همکاران [۸۱] معرفی شد. هم‌چنین ما و شانبلات^{۳۴} [۶۰] یک شبکه عصبی دو فازی ارائه کردند که فاز اول شبکه مشابه شبکه کندی و چا بود و در فاز دوم آن مسیر شبکه به جواب دقیق مسأله همگرا می‌شد بنابراین این روش جواب‌های دقیق‌تری نسبت به شبکه کندی و چا ارائه می‌داد. مشکل این شبکه در این بود که پایداری فاز دوم شبکه بستگی به انتخاب یک مقدار بزرگ از پارامتر جریمه داشت و لذا اگرچه تأثیر پارامتر جریمه در این روش کاهش یافته بود و جواب‌های حاصل دقیق‌تر بودند اما هنوز مستقل از پارامتر نبود.

ژانگ و کنستانتینیدس^{۳۵} [۱۱۳] بر مبنای روش لاگرانژ یک شبکه عصبی ارائه کردند که کاملاً مستقل از پارامتر جریمه و قادر به حل مسائل غیرخطی بود. نقطه تعادل این شبکه عصبی در شرایط بهینگی مرتبه اول و دوم صدق می‌کرد و هم‌چنین شبکه حاصل همگرا بود. در سال ۱۹۹۳ بوزردوم و پتیسن^{۳۶} [۱۳] شبکه‌ای را بر مبنای گرادیان و روش تصویر و مستقل از پارامتر جریمه ابداع کردند که تنها قادر به حل مسائل درجه دوم با متغیرهای کران‌دار بود. این روش در عمل روش کارایی بود اما نمی‌توانست مسائل کلی برنامه‌ریزی خطی و درجه دوم را حل کند.

زیا^{۳۷} و همکاران [۱۰۱، ۱۰۲، ۱۰۳، ۱۰۴، ۱۰۵، ۱۰۶] چندین مدل را برای حل مسائل بهینه‌سازی غیرخطی با قيود خطی و غیرخطی ارائه کردند. هم‌چنین چندین مدل شبکه عصبی برای حل مسائل بهینه‌سازی غیرخطی توسط عفتی و همکاران [۲۰، ۲۱، ۲۲، ۶۵، ۶۶] ارائه شده است.

۲.۱ مقدمه‌ای بر بهینه‌سازی غیرخطی

تعریف ۱.۲.۱. [۵۹] تابع $f: \mathbb{R}^n \rightarrow \mathbb{R}$ روی مجموعه محدب و ناتهی $S \subseteq \mathbb{R}^n$ محدب نامیده می‌شود اگر برای هر $x, y \in S$ و $\alpha \in (0, 1)$ داشته باشیم

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$$

^{۳۲} Switched-Capacitor

^{۳۳} Rodriguez-Vazquez

^{۳۴} Maa and Shanblatt

^{۳۵} Zhang and Constantinides

^{۳۶} Bouzerdoum and Pattison

^{۳۷} Xia

تابع f روی S به طور اکید محدب نامیده می‌شود اگر نامساوی بالا زمانی که $x \neq y$ است، به‌طور اکید برقرار باشد.

یک مسأله برنامه‌ریزی غیرخطی در حالت کلی به‌صورت زیر در نظر گرفته می‌شود

$$\text{minimize } f(x) \quad (1.1)$$

$$\text{subject to } g_i(x) \leq 0, \quad (2.1)$$

$$h_j(x) = 0, \quad (3.1)$$

$$x \in X \subset \mathbb{R}^n, \quad (4.1)$$

که $f: \mathbb{R}^n \rightarrow \mathbb{R}$ ، $g_i: \mathbb{R}^n \rightarrow \mathbb{R}$ ($i = 1, \dots, m$) و $h_j: \mathbb{R}^n \rightarrow \mathbb{R}$ ($j = 1, \dots, p$) توابع دو بار مشتق پذیر پیوسته و X زیر مجموعه‌ای باز از \mathbb{R}^n است. برای تسهیل در نمادگذاری، توابع برداری $g = (g_1, g_2, \dots, g_m)^T$ و $h = (h_1, h_2, \dots, h_p)^T$ را معرفی می‌کنیم و (۱.۱)–(۴.۱) را به‌صورت زیر می‌نویسیم

$$\text{minimize } f(x) \quad (5.1)$$

$$\text{subject to } g(x) \leq 0, \quad (6.1)$$

$$h(x) = 0, \quad (7.1)$$

$$x \in X \subset \mathbb{R}^n, \quad (8.1)$$

که محدودیت‌های $g(x) \leq 0$ و $h(x) = 0$ ، محدودیت‌های تابعی و محدودیت $x \in X \subset \mathbb{R}^n$ محدودیت مجموعه‌ای نامیده می‌شود. هرگاه توابع برداری f و g روی X محدب باشند و h آفین باشد یعنی $h = Ax - b$ ، $A \in \mathbb{R}^{p \times n}$ ، $rank(A) = p$ ($0 \leq p < n$) و $b \in \mathbb{R}^p$ باشد، آن‌گاه (۵.۱)–(۸.۱) یک مسأله بهینه‌سازی محدب^{۳۸} و در غیر این صورت یک مسأله بهینه‌سازی نامحدب^{۳۹} نامیده می‌شود.

تعریف ۲.۲.۱. [۵۹] نقطه x جواب شدنی^{۴۰} مسأله (۵.۱)–(۸.۱) نامیده می‌شود هرگاه در محدودیت‌های مسأله صدق کند. نقطه شدنی x یک نقطه منظم^{۴۱} نامیده می‌شود اگر بردارهای گرادیان $\nabla g_i(x)$ ، $i \in \{i | g_i(x) = 0\}$ و $\nabla h_j(x)$ ، $j = 1, \dots, p$ مستقل خطی باشند.

قضیه ۱.۲.۱. (شرایط کافی مرتبه دوم^{۴۲}) [۵۹] فرض کنید x^* یک نقطه شدنی برای مسأله

^{۳۸}Convex optimization

^{۳۹}Nonconvex optimization

^{۴۰}Feasible solution

^{۴۱}Regular point

^{۴۲}Second-order sufficient conditions

(۵.۱)–(۸.۱) باشد. اگر $\mu \in \mathbb{R}^m$ و $\lambda \in \mathbb{R}^p$ وجود داشته باشند به طوری که

$$\mu \geq \circ, \quad (۹.۱)$$

$$\mu^T \mathbf{g}(x^*) = \circ, \quad (۱۰.۱)$$

$$\nabla f(x^*) + \lambda^T \nabla \mathbf{h}(x^*) + \mu^T \nabla \mathbf{g}(x^*) = \circ, \quad (۱۱.۱)$$

و ماتریس هسین^{۴۳}

$$\nabla^2 L(x^*, \lambda, \mu) = \nabla^2 f(x^*) + \sum_{i=1}^m \lambda_i \nabla^2 g_i(x^*) + \sum_{j=1}^p \mu_j \nabla^2 h_j(x^*) \quad (۱۲.۱)$$

بر زیرفضای

$$M(x^*) = \{d \mid \nabla \mathbf{h}(x^*)d = \circ, \nabla g_j(x^*)d = \circ, \forall j \in J(x^*)\},$$

که در آن

$$J(x^*) = \{j \mid g_j(x^*) = \circ, \mu_j > \circ\},$$

معین مثبت باشد، آن گاه x^* یک نقطه مینیمم محلی اکید^{۴۴} برای مسأله (۵.۱)–(۸.۱) است.

مسأله دوگان لاگرانژ

مسأله برنامه‌ریزی غیرخطی (۵.۱)–(۸.۱) را در نظر بگیرید، این مسأله را مسأله اولیه (P) نیز می‌نامیم. مسأله دوگان لاگرانژ^{۴۵} (D) به صورت زیر تعریف می‌شود

$$\max \quad \theta(u, v) \quad (۱۳.۱)$$

$$\text{subject to } u \geq \circ, \quad (۱۴.۱)$$

که در آن

$$\theta(u, v) = \inf_x \left\{ f(x) + \sum_{i=1}^m u_i g_i(x) + \sum_{j=1}^p v_j h_j(x), x \in X \right\}, \quad (۱۵.۱)$$

$$u = (u_1, u_2, \dots, u_m)^T, \quad v = (v_1, v_2, \dots, v_p)^T,$$

u_i – امین مؤلفه از u ، متغیر دوگان^{۴۶} یا ضرب‌گر لاگرانژ^{۴۷} مربوط به قید $g_i(x) \leq \circ$ می‌باشد و v_j – امین مؤلفه v ، متغیر دوگان یا ضرب‌گر لاگرانژ مربوط به قید $h_j(x) = \circ$ می‌باشد.

^{۴۳} Hessian matrix

^{۴۴} Strict local minimum point

^{۴۵} Lagrangian dual problem

^{۴۶} Dual variable

^{۴۷} Lagrangian multiplier

مسأله دوگان لاگرانژ (۱۳.۱) و (۱۴.۱) در شکل برداری به صورت زیر نوشته می شود

$$\text{maximize } \theta(u, v) \quad (۱۶.۱)$$

$$\text{subject to } u \geq 0, \quad (۱۷.۱)$$

که در آن

$$\theta(u, v) = \inf_x \{f(x) + u^t g(x) + v^t h(x), x \in X\},$$

هم چنین $h = (h_1, \dots, h_p) : \mathbb{R}^n \rightarrow \mathbb{R}^p$ و $g = (g_1, \dots, g_m)^T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ، $f : \mathbb{R}^n \rightarrow \mathbb{R}$ هستند.

قضیه ۲.۲.۱. (دوگانی موضعی)^{۴۸} [۵۹] فرض کنید مسأله

$$\text{minimize } f(x)$$

$$\text{subject to } h(x) = 0,$$

$$x \in X \subset \mathbb{R}^n,$$

که $f, h \in C^2$ و $h(x) \in \mathbb{R}^m$ دارای یک جواب موضعی در x^* با مقدار متناظر r^* و ضریب لاگرانژ v^* است. هم چنین فرض کنید که x^* یک نقطه منظم برای قیود است و هسین لاگرانژ متناظر $\nabla^2 L(x^*, v^*)$ معین مثبت است. در این صورت مسأله دوگان

$$\text{max } \theta(v) \quad (۱۸.۱)$$

که در آن

$$\theta(v) = \inf_x \{f(x) + v^t h(x), x \in X\},$$

دارای یک جواب موضعی در v^* با مقدار متناظر r^* است و x^* جواب متناظر با v^* در تعریف θ است.

به سادگی می توان نتایج فوق را به مسائلی که علاوه بر قیود تساوی قیود نامساوی هم دارند تعمیم داد [۵۹].

۳.۱ سیستم های دینامیکی

تعریف ۱.۳.۱. [۷۶] سیستم دینامیکی^{۴۹} زیر را در نظر بگیرید:

$$\dot{x} = f(x(t)), x(t_0) = x_0 \in \mathbb{R}^n, \quad (۱۹.۱)$$

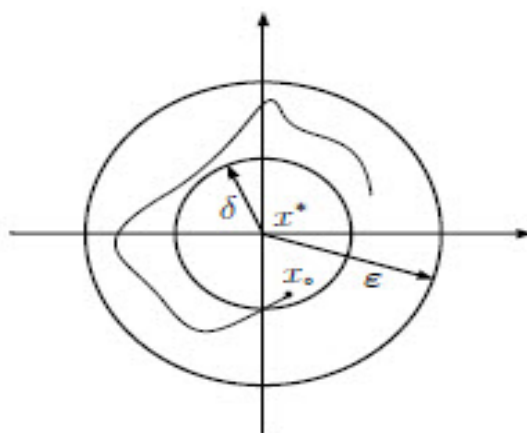
^{۴۸}Local duality

^{۴۹}Dynamic system

که در آن f یک تابع برداری از \mathbb{R}^n به \mathbb{R}^n می‌باشد. x^* یک نقطه تعادل $^{\circ 5}$ (۱۹.۱) نامیده می‌شود اگر $f(x^*) = 0$. اگر همسایگی $\Omega^* \subseteq \mathbb{R}^n$ از x^* وجود داشته باشد که $f(x^*) = 0$ و $\forall x \in \Omega^* \setminus \{x^*\}$ ، $f(x) \neq 0$ ، آن‌گاه x^* یک نقطه تعادل تنها $^{\circ 51}$ نامیده می‌شود.

تعریف ۲.۳.۱. (پایداری به مفهوم لیاپانوف) $^{\circ 52}$ [۷۶] فرض می‌کنیم که $x(t)$ یک جواب (۱۹.۱) باشد، نقطه تعادل تنها x^* ، پایدار به مفهوم لیاپانوف است اگر برای هر $x_0 = x(t_0)$ و هر $\varepsilon > 0$ ، یک $\delta > 0$ وجود داشته باشد به طوری که اگر $\|x(t_0) - x^*\| < \delta$ آن‌گاه

$$\|x(t) - x^*\| < \varepsilon, \forall t \geq t_0.$$



شکل ۵.۱: پایداری لیاپانوف.

تعریف ۳.۳.۱. [۷۶] سیستم دینامیکی (۱۹.۱) همگرای سراسری $^{\circ 53}$ به مجموعه جواب‌های (۱۹.۱) که با Ω^* نمایش داده می‌شود، گفته می‌شود اگر برای هر نقطه آغازین دلخواه، مسیر $x(t)$ در رابطه زیر صدق کند:

$$\lim_{t \rightarrow \infty} \text{dist}(x(t), \Omega^*) = 0,$$

که در آن

$$\text{dist}(x, \Omega^*) = \inf_{y \in \Omega^*} \|x - y\|,$$

و $\|\cdot\|$ نرم l_2 را نشان می‌دهد.

تعریف ۴.۳.۱. [۷۶] سیستم دینامیکی (۱۹.۱) در نقطه تعادل x^* که یکتا می‌باشد، پایدار مجانبی سراسری $^{\circ 54}$ نامیده می‌شود اگر x^* به مفهوم لیاپانوف پایدار باشد و در شرط صدق

$^{\circ 50}$ Equilibrium point

$^{\circ 51}$ Isolated equilibrium point

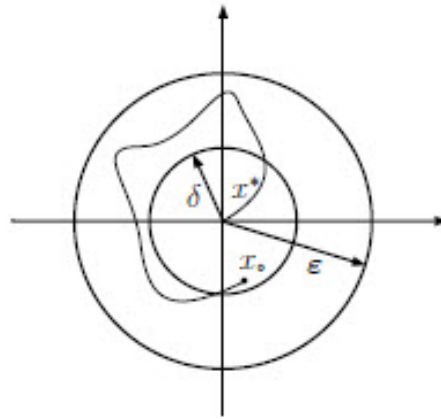
$^{\circ 52}$ Stability in the sense of Lyapunov

$^{\circ 53}$ Globally convergent

$^{\circ 54}$ Globally asymptotically stable

کند:

$$\lim_{t \rightarrow \infty} x(t) = x^*.$$



شکل ۶.۱: پایداری مجانبی.

تعریف ۵.۳.۱. [۷۶] سیستم دینامیکی (۱۹.۱) در نقطه x^* پایداری نمایی سراسری^{۵۵} با نرخ β نامیده می‌شود اگر برای هر نقطه آغازین دلخواه، مسیر $x(t)$ در رابطه زیر صدق کند

$$\|x(t) - x^*\| \leq \alpha \|x(t_0) - x^*\| e^{-\beta(t-t_0)}, \quad \forall t \geq t_0,$$

که در آن α و β ثابت‌های مثبت و مستقل از نقاط آغازین هستند. واضح است که پایداری نمایی سراسری پایداری مجانبی سراسری را نتیجه می‌دهد.

تعریف ۶.۳.۱. [۷۶] مجموعه $M \subseteq \mathbb{R}^n$ یک مجموعه تغییر ناپذیر^{۵۶} نسبت به سیستم (۱۹.۱) گفته می‌شود اگر $x(t_0) \in M$ و $t_0 \geq 0$ ، به ازای هر $t \geq t_0$ ، $x(t) \in M$ باشد.

تعریف ۷.۳.۱. [۷۶] فرض می‌کنیم که $\Omega \subset \mathbb{R}^l$ یک مجموعه بسته و محدب باشد که ممکن است بیکران باشد. $P_\Omega(x): \mathbb{R}^l \rightarrow \Omega$ یک عملگر تصویر^{۵۷} نامیده می‌شود و به صورت زیر تعریف می‌شود

$$P_\Omega(x) = \arg \min_{v \in \Omega} \|x - v\|, \quad (۲۰.۱)$$

که $\|\cdot\|$ نرم l_2 را نشان می‌دهد.

^{۵۵} Globally exponential stability

^{۵۶} Invariant set

^{۵۷} Projection operator

تعریف ۸.۳.۱. [۷۶] تابع $F : E \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ روی E در شرط لیب شیتز صدق می‌کند (پیوسته لیب شیتز^{۵۸} است) اگر ثابت L وجود داشته باشد به طوری که برای هر دو نقطه $x, y \in E$ داشته باشیم:

$$\|F(x) - F(y)\| \leq L \|x - y\|.$$

تابع F پیوسته لیب شیتز محلی^{۵۹} روی E نامیده می‌شود اگر برای نقطه $x_0 \in E$ ، یک $\varepsilon > 0$ همسایگی مانند $N_\varepsilon(x_0) \subset E$ و یک ثابت L_0 وجود داشته باشد به طوری که برای هر دو نقطه $x, y \in N_\varepsilon(x_0)$ داشته باشیم:

$$\|F(x) - F(y)\| \leq L_0 \|x - y\|.$$

قضیه ۱.۳.۱. [۷۶] تابع $F : E \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ را در نظر بگیرید. اگر $F \in C^1(E)$ ، آن‌گاه F روی E لیب شیتز محلی است.

تعریف ۹.۳.۱. نگاشت $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$

(الف) یکنوا^{۶۰} نامیده می‌شود اگر برای هر دو نقطه $x, y \in \mathbb{R}^n$ داشته باشیم

$$\langle x - y, F(x) - F(y) \rangle \geq 0.$$

(ب) اکیداً یکنوا^{۶۱} نامیده می‌شود اگر برای هر دو نقطه متمایز $x, y \in \mathbb{R}^n$ داشته باشیم

$$\langle x - y, F(x) - F(y) \rangle > 0.$$

(پ) قویاً یکنوا^{۶۲} نامیده می‌شود اگر یک ثابت $\beta > 0$ وجود داشته باشد که برای هر دو نقطه $x, y \in \mathbb{R}^n$ داشته باشیم

$$\langle x - y, F(x) - F(y) \rangle \geq \beta \|x - y\|^2.$$

که $\langle \cdot, \cdot \rangle$ ضرب داخلی را نشان می‌دهد.

تعریف ۱۰.۳.۱. [۷۳] اگر $g : \Omega_1 \subset \mathbb{R}^l \rightarrow \mathbb{R}$ ، آن‌گاه هر مجموعه غیر تهی به شکل زیر

$$L(r) = \{u \in \Omega_1 \mid g(u) \leq r, r \in \mathbb{R}\}$$

یک مجموعه سطح^{۶۳} از g نامیده می‌شود.

^{۵۸}Lipschitz continuous

^{۵۹}Locally lipschitz continuous

^{۶۰}Monotone map

^{۶۱}Strictly monotone map

^{۶۲}Strongly monotone map

^{۶۳}Level set

تعریف ۱۱.۳.۱ [۷۳] یک ماتریس $n \times n$ ، $M(x)$ که عناصر آن m_{ij} و $i = 1, \dots, n$ و $j = 1, \dots, n$ توابعی هستند که روی مجموعه $S \subset \mathbb{R}^n$ تعریف شده‌اند، روی S نیمه معین مثبت^{۶۴} نامیده می‌شود اگر

$$v^T M(x) v \geq 0, \forall v \in \mathbb{R}^n, \forall x \in S.$$

ماتریس $M(x)$ روی S معین مثبت قوی^{۶۵} نامیده می‌شود اگر عددی مثبت مانند α وجود داشته باشد به قسمی که:

$$v^T M(x) v \geq \alpha \|v\|^2, \forall v \in \mathbb{R}^n, \forall x \in S.$$

اگر $\gamma(x)$ کوچکترین مقدار ویژه ماتریس $M(x)$ باشد سه حالت زیر را می‌توان نتیجه گرفت:

۱. $M(x)$ روی S نیمه معین مثبت است اگر و فقط اگر به ازای هر $x \in S$ ، $\gamma(x) \geq 0$.

۲. $M(x)$ روی S معین مثبت^{۶۶} است اگر و فقط اگر به ازای هر $x \in S$ ، $\gamma(x) > 0$.

۳. $M(x)$ روی S قویاً معین مثبت است اگر و فقط اگر به ازای هر $x \in S$ ، $\gamma(x) \geq \alpha > 0$.

تعریف ۱۲.۳.۱ [۷۳] فرض می‌کنیم $V: \mathbb{R}^n \rightarrow \mathbb{R}$ یک تابع به‌طور پیوسته مشتق پذیر باشد. $V(x)$ به‌طور شعاعی بی‌کران گفته می‌شود اگر

$$\|x\| \rightarrow \infty \Rightarrow V(x) \rightarrow \infty.$$

لم ۱.۳.۱ [۷۳] فرض کنیم که $g: \Omega_1 \subset \mathbb{R}^l \rightarrow \mathbb{R}$ باشد که در آن Ω_1 بی‌کران است. آن‌گاه همه مجموعه‌های سطح g کراندار می‌باشند اگر و فقط اگر

$$\lim_{k \rightarrow \infty} g(u^k) = +\infty,$$

که در آن $\{u^k\} \subset \Omega_1$ و $\lim_{k \rightarrow \infty} \|u^k\| = +\infty$.

لم ۲.۳.۱ [۷۳] فرض کنیم که $g: \Omega_1 \subset \mathbb{R}^l \rightarrow \mathbb{R}$ روی مجموعه بسته Ω_1 پیوسته باشد آن‌گاه همه مجموعه‌های سطح g کراندار می‌باشند اگر و فقط اگر مجموعه مینیمم‌کننده‌های g غیرتهی و کراندار باشند.

قضیه ۲.۳.۱ (قضیه نقطه ثابت بروئر^{۶۷}) [۷۳] تابع پیوسته $f: \Omega \rightarrow \Omega$ ، که Ω یک مجموعه فشرده و محدب می‌باشد حداقل یک نقطه ثابت دارد.

^{۶۴}Positive semidefinite

^{۶۵}Strongly positive definite

^{۶۶}Positive definite

^{۶۷}Brouwer's fixed point theorem

قضیه ۳.۳.۱. [۷۳] فرض کنیم که تابع $F : K \subset \mathbb{R}^l \rightarrow \mathbb{R}^l$ روی K به‌طور پیوسته مشتق‌پذیر باشد، آن‌گاه F روی K یکنوا (اکیداً یکنوا) است اگر و فقط اگر ماتریس ژاکوبین آن $\nabla F(x)$ برای هر $x \in K$ نیمه‌معین مثبت (معین مثبت) باشد.

قضیه ۴.۳.۱. [۷۳] فرض کنیم که تابع $F(x)$ روی K به‌طور پیوسته مشتق‌پذیر باشد و ژاکوبین $F(x)$ قویاً معین مثبت باشد آن‌گاه $F(x)$ قویاً یکنوا می‌باشد.

قضیه ۵.۳.۱. [۷۶] در سیستم دینامیکی (۱۹.۱) فرض کنیم که f یک تابع پیوسته از \mathbb{R}^n به \mathbb{R}^n باشد، آن‌گاه برای هر $t_0 \geq 0$ و $x_0 \in \mathbb{R}^n$ ، یک جواب محلی $x(t)$ به ازای $t \in [t_0, \tau)$ که $\tau > t_0$ وجود دارد. علاوه بر این اگر f در x_0 پیوسته لیپ‌شیتز محلی باشد آن‌گاه جواب یکتاست و اگر f در \mathbb{R}^n پیوسته لیپ‌شیتز باشد آن‌گاه τ می‌تواند تا $+\infty$ توسعه داده شود.

تعریف ۱۳.۳.۱. [۷۶] اگر یک جواب محلی تعریف شده در بازه $[t_0, \tau)$ نتواند به یک جواب محلی روی یک بازه بزرگتر $[t_0, \tau_1)$ که $\tau_1 > \tau$ می‌باشد گسترش یابد، آن‌گاه یک جواب ماکسیمال نامیده می‌شود و بازه $[t_0, \tau)$ ، بازه ماکسیمال وجود جواب نامیده می‌شود. بازه ماکسیمال وجود جواب مربوط به x_0 اغلب به‌صورت $[t_0, \tau(x_0))$ تعریف می‌شود.

قضیه ۶.۳.۱. [۱۱۲] فرض کنیم $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ یک تابع پیوسته باشد. اگر $x(t) \in [t_0, \tau(x_0))$ یک جواب ماکسیمال باشد و $\tau(x_0) < +\infty$ آن‌گاه

$$\lim_{t \rightarrow \tau(x_0)} \|x(t)\| = +\infty.$$

قضیه ۷.۳.۱. (قضیه پایداری لیاپانوف) [۵۰] فرض کنید که $x = 0$ یک نقطه تعادل سیستم (۱۹.۱) باشد و $V : \mathbb{R}^n \rightarrow \mathbb{R}$ یک تابع به‌طور پیوسته مشتق‌پذیر باشد به‌طوری‌که:

$$1. \quad V(0) = 0$$

$$2. \quad \text{به ازای هر } x \in \mathbb{R}^n \setminus \{0\}, \quad V(x) > 0$$

$$3. \quad \text{به ازای هر } x \in \mathbb{R}^n \setminus \{0\}, \quad \dot{V}(x) \leq 0$$

آن‌گاه $x = 0$ نقطه پایداری سیستم خواهد بود.

قضیه ۸.۳.۱. (قضیه پایداری جانبی) [۵۰] تحت شرایط قضیه ۷.۳.۱ اگر $V(\cdot)$ در شرایط زیر صدق کند

$$1. \quad V(0) = 0$$

$$2. \quad \text{به ازای هر } x \in \mathbb{R}^n \setminus \{0\}, \quad V(x) > 0$$

$$3. \quad \text{به ازای هر } x \in \mathbb{R}^n \setminus \{0\}, \quad \dot{V}(x) < 0$$

آن گاه $x = 0$ پایدار مجانبی خواهد بود.

قضیه ۹.۳.۱. (قضیه پایداری مجانبی سراسری) [۵۰] تحت شرایط قضیه ۷.۳.۱ اگر $V(\cdot)$ به طور شعاعی بی کران باشد آن گاه نقطه $x = 0$ پایدار مجانبی سراسری است.

قضیه ۱۰.۳.۱. (اصل تغییرناپذیری لاسال^{۶۸}) [۵۰] فرض کنیم یک تابع به طور پیوسته مشتق پذیر باشد، همچنین فرض کنیم

۱. $M \subset \mathbb{R}^n$ یک مجموعه فشرده و پایدار نسبت به جواب سیستم (۱۹.۱) باشد.

۲. به ازای $x \in M$ ، $\dot{V}(x) \leq 0$.

۳. E مجموعه همه نقاط M باشد به طوری که $\dot{V}(x) = 0$.

۴. N بزرگترین مجموعه پایدار در E باشد.

آن گاه به ازای هر $x(t_0) \in M$ که $t_0 \geq 0$ می باشد داریم:

$$\lim_{t \rightarrow \infty} \text{dist}(x(t), N) = 0.$$

۴.۱ نامساوی های وردشی

مقدمه ای بر نامساوی وردشی

نظریه نامساوی وردشی^{۶۹} اولین بار در سال ۱۹۶۶ به وسیله هارتمن^{۷۰} و استمپکیا^{۷۱} به عنوان ابزاری برای مطالعه معادلات دیفرانسیل جزئی با کاربردهایی در مکانیک معرفی شد [۳۱]. این نامساوی ها با بعد نامتناهی بودند. نظریه نامساوی وردشی با بعد متناهی در سال ۱۹۸۰ که به وسیله دافرmos^{۷۲} معرفی شد، دافرmos متوجه شد که شرایط تعادل شبکه ترافیک که به وسیله اسمیت^{۷۳} در سال ۱۹۷۹ بیان شد ساختار نامساوی وردشی دارد. نامساوی های وردشی در اقتصاد، علم مدیریت، تحقیق در عملیات و همچنین در مهندسی کاربردهای فراوانی دارد.

تعریف ۱.۴.۱. فرض کنید $F : K \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ تابعی پیوسته روی مجموعه بسته و محدب K باشد، مسأله نامساوی وردشی که آن را با نماد $VI(F, K)$ نشان می دهند یافتن یک بردار $x^* \in K$ است به طوری که

$$F(x^*)^T(x - x^*) \geq 0, \quad \forall x \in K, \quad (21.1)$$

^{۶۸} LaSalle principle of invariance

^{۶۹} Variational inequality

^{۷۰} Hartman

^{۷۱} Stampacchia

^{۷۲} Dafermos

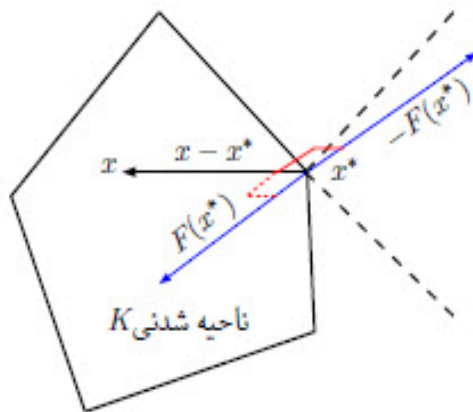
^{۷۳} Smith

یا به‌طور معادل

$$\langle F(x^*), x - x^* \rangle \geq 0, \quad \forall x \in K, \quad (22.1)$$

که در آن $\langle \cdot, \cdot \rangle$ ضرب داخلی در فضای اقلیدسی n بعدی است.

از دیدگاه هندسی، نامساوی وردشی (۲۲.۱) بیان می‌کند که $F(x^*)$ بر ناحیه شدنی K در نقطه x^* عمود است (شکل ۷.۱).



شکل ۷.۱: تعبیر هندسی مسأله نامساوی وردشی.

قضیه ۱.۴.۱. [۹] مسأله بهینه‌سازی زیر را در نظر بگیرید:

$$\begin{aligned} & \text{minimize } f(x) \\ & \text{subject to } x \in K, \end{aligned}$$

که در آن f یک تابع محدب و به‌طور پیوسته مشتق پذیر و K یک مجموعه محدب و غیرتهی در \mathbb{R}^n است آن‌گاه x^* جواب بهینه مسأله فوق است اگر و تنها اگر x^* جواب مسأله نامساوی وردشی

$$\nabla f(x^*)^T (x - x^*) \geq 0, \quad \forall x \in K$$

باشد.

قضیه ۲.۴.۱. [۲] فرض کنید تابع $F : K \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ روی K به‌طور پیوسته مشتق پذیر باشد و ماتریس ژاکوبین F متقارن و نیمه معین مثبت باشد، آن‌گاه تابع محدب حقیقی مقدار $f : K \rightarrow \mathbb{R}$ وجود دارد که در رابطه

$$\nabla f(x) = F(x),$$

صدق می‌کند و x^* جواب نامساوی وردشی $VI(F, K)$ است اگر و فقط اگر x^* جواب بهینه مسأله زیر باشد:

$$\begin{aligned} & \text{minimize } f(x) \\ & \text{subject to } x \in K, \end{aligned}$$

تعریف ۲.۴.۱. [۲] فرض می‌کنیم \mathbb{R}_+^n ناحیه نامنفی روی \mathbb{R}^n را نشان دهد و فرض کنیم $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$ باشد. مسأله مکمل غیرخطی^{۷۴} روی \mathbb{R}_+^n یک سیستم از معادلات و نامعادلات است به صورت زیر است:
یافتن $x^* \geq 0$ به طوری که

$$F(x^*) \geq 0, \quad F(x^*)^T x^* = 0. \quad (23.1)$$

هرگاه نگاشت F آفین باشد یعنی $F(x) = Mx + b$ که در آن M یک ماتریس $n \times n$ و b یک بردار $n \times 1$ می‌باشد مسأله (۲۳.۱) یک مسأله مکمل خطی^{۷۵} نامیده می‌شود.

قضیه زیر رابطه بین مسأله مکمل و مسأله نامساوی وردشی را نشان می‌دهد.

قضیه ۳.۴.۱. [۲] مسأله نامساوی وردشی $VI(F, \mathbb{R}_+^n)$ و مسأله مکمل (۲۳.۱) جواب‌های یکسان دارند.

^{۷۴}Nonlinear complementarity problem (NCP)

^{۷۵}Linear complementarity problem (LCP)

فصل ۲

مقدمه ای بر مسائل بهینه سازی نیمه معین

۱.۲ مقدمه

در مسائل برنامه‌ریزی نیمه‌معین^۱ شخص یک تابع خطی را به شرط محدودیتی که ترکیبی آفین از ماتریس‌های متقارن نیمه‌معین مثبت^۲ است، مینیمم می‌کند. این چنین محدودیتی غیرخطی و محدب است. بنابراین برنامه‌ریزی‌های نیمه‌معین مسائل بهینه‌سازی محدب هستند. برنامه‌ریزی نیمه‌معین چندین مساله استاندارد را یک شکل می‌کند (برای مثال برنامه‌ریزی خطی و درجه دوم) و کاربردهای بسیاری در مهندسی و بهینه‌سازی ترکیبیاتی پیدا می‌کند. اگر چه برنامه‌ریزی‌های نیمه‌معین کلی‌تر از برنامه‌ریزی‌های خطی هستند، حل آن‌ها سخت‌تر نیست. بیشتر روش‌های نقطه درونی برای برنامه‌ریزی خطی برای برنامه‌ریزی نیمه‌معین تعمیم داده شده‌اند.

یکی از اولین مقالات در زمینه خواص نظری مسائل نیمه معین توسط بلمن^۳ و فن^۴ [۱۰]

^۱ Semidefinite Programming(SDP)

^۲ Positive semidefinite maTrix

^۳ Bellman

^۴ Fan

ارائه شد. شرایط بهینگی مسائل نیمه معین را در [۱۶، ۲۹، ۵۴] می توان یافت. بسیاری از محققان روی مسئله مینیم کردن بزرگترین مقدار ویژه ماتریس متقارن کار کرده اند که می تواند به عنوان یک مسئله نیمه معین بیان شود، برای مثال [۱۸، ۲۵، ۷۵] را ببینید. روش های نقطه درونی برای مسائل برنامه ریزی خطی توسط کارماکار^۵ در ۱۹۸۴ [۴۷] معرفی شدند. نسترو^۶ و نیمروفسکی^۷ در [۷۲] نشان دادند که روش های نقطه درونی می توانند برای حل مسائل بهینه سازی محدب توسعه یابند. از آنجایی که مسائل برنامه ریزی نیمه معین دسته مهمی از مسائل بهینه سازی محدب هستند، روش های نقطه درونی برای حل آنها نیز به کار گرفته شدند. علیزاده^۸ [۵] و کاماس^۹ و کارماکار [۴۵]، مستقل از نسترو و نیمروفسکی، روش های نقطه درونی را از مسائل خطی به مسائل نیمه معین توسعه دادند. مقالات دیگری نیز در این زمینه توسط جیر^{۱۰} [۴۳]، واندربرگ^{۱۱} و بوید^{۱۲} [۸۹] ارائه شده است.

۲.۲ تعاریف و نمادها

برای تعریف و بررسی خواص برنامه ریزی های نیمه معین نیاز به تعاریفی از جبرخطی داریم. در سرتاسر این رساله ما از نمادها و تعاریف زیر استفاده خواهیم کرد:

- ماتریس متقارن $A \in \mathbb{R}^{n \times n}$ را نیمه معین مثبت (نیمه معین منفی) می گویند هرگاه به ازای هر $x \in \mathbb{R}^n$ داشته باشیم $x^T A x \geq 0$ (همچنین A را معین مثبت (معین منفی) گویند هرگاه برای هر $x \neq 0$ داشته باشیم $x^T A x > 0$ (نیمه معین مثبت (نیمه معین منفی) را با نماد $A \succeq 0$ ($A \preceq 0$) نمایش می دهند. همچنین اگر A معین مثبت (معین منفی) باشد، با نماد $A \succ 0$ ($A \prec 0$) نشان داده می شود.
- مجموعه تمام ماتریس های حقیقی $m \times n$ را با $M_{m,n}$ ، تمام ماتریس های متقارن $m \times m$ بعدی را با S^m و تمام ماتریس های متقارن نیمه معین مثبت (نیمه معین منفی) را با S^m_+ (S^m_-) نمایش می دهیم. برای تمام این مجموعه از ماتریس ها از حاصل ضرب داخلی یا تریس^{۱۳}

$$\langle A, B \rangle = Tr(B^T A) = Tr(A^T B) = \sum_{i=1}^m \sum_{j=1}^m a_{ij} b_{ji},$$

^۵Karmakar

^۶Nesterov

^۷Nemirovsky

^۸Alizadeh

^۹kamath

^{۱۰}Jarre

^{۱۱}Vandenberghe

^{۱۲}Boyd

^{۱۳}Trace product

و نرم فریبینیوس^{۱۴} $\|A\| = \sqrt{\langle A, A \rangle}$ استفاده خواهیم کرد. خواص زیر از ماتریس‌های متقارن را یادآوری می‌کنیم.

- ۱- اگر $A \in S^n$ ، آنگاه ماتریس متعامد P و ماتریس قطری یا بلوکی قطری D وجود دارند به طوری که $A = P^T D P$.
- ۲- اگر $A = P^T D P$ ، آنگاه ستون‌های P بردارهای ویژه‌ی متناظر با مقادیر ویژه‌ی ماتریس A هستند که عناصر ماتریس D را تشکیل می‌دهند.
- ۳- $A \succeq 0$ اگر و فقط اگر $A = P^T D P$ که در آن مقادیر ویژه (یعنی عناصر ماتریس قطری یا بلوکی قطری D) همگی نامنفی هستند.
- ۴- $A \succ 0$ اگر و فقط اگر $A = P^T D P$ که در آن مقادیر ویژه (یعنی عناصر ماتریس قطری یا بلوکی D) همگی مثبت هستند.
- ۵- (لم مکمل شور) فرض کنید ماتریس M به صورت زیر تعریف شده باشد

$$M = \begin{pmatrix} P & v \\ v^T & d \end{pmatrix}$$

که در آن $P \succ 0$ ، v یک بردار و d یک اسکالر است. آنگاه

$$M \succ 0 \iff d - v^T P^{-1} v \geq 0. \quad (1.2)$$

- برای ماتریس $A \in M_{m,m}$ ، $[A]_+$ را تصویر عمودی روی S_+^m را به صورت زیر تعریف می‌کنند، یعنی

$$[A]_+ = \operatorname{argmin}_{B \in S_+^m} \|A - B\|.$$

- عملگر $\operatorname{vec}(\cdot) : M_{m,n} \rightarrow \mathbb{R}^{mn}$ برای تبدیل ماتریس‌ها به بردارها استفاده خواهد شد، به این ترتیب که اگر $A = (a_{ij}) \in M_{m,n}$ آنگاه $\operatorname{vec}(A) = (a_{11}, a_{21}, \dots, a_{m1}, a_{12}, \dots, a_{mn})^T$.
- برای تابع ماتریس مقدار $A(\cdot)$ که روی \mathbb{R}^n تعریف شده است از نماد زیر برای عملگر مشتق در x_* استفاده می‌کنیم

$$DA(x_*) = \left(\frac{\partial A(x_*)}{\partial x_1}, \dots, \frac{\partial A(x_*)}{\partial x_n} \right)^T, \quad (2.2)$$

- اگر عملگر خطی V از \mathbb{R}^n به S^m را با $Vy = \sum_{i=1}^n y_i V_i$ تعریف کنیم که $V_i \in S^m$ برای تمام $i \in \{1, \dots, n\}$ ، عملگر V^* با رابطه زیر بیان می‌شود

$$V^* Z = (\langle V_1, Z \rangle, \dots, \langle V_n, Z \rangle)^T, \quad \forall V \in S^m. \quad (3.2)$$

^{۱۴}Frobenius norm

لم ۱.۲.۲. [۸۷] برای هر $A \in S^m$ عبارت های زیر معادلند:

الف. $A \in S_+^m$.

ب. $\lambda_j \geq 0$ ، $(j = 1, \dots, m)$ ، که نشان دهنده j -امین مقدار ویژه ماتریس A است.

پ. $\langle A, B \rangle \geq 0$ ، برای تمام $B \in S_+^m$.

لم ۲.۲.۲. [۸۷]

الف. برای هر $A \in S^m$ داریم

$$[A]_+ := P \text{diag}((\lambda_1)^+, \dots, (\lambda_m)^+) P^T,$$

که $(\lambda)^+ = \max\{0, \lambda\}$ ، نشان دهنده j -امین مقدار ویژه ماتریس A و P ماتریسی

است که در تجزیه ماتریس A ظاهر می شود، یعنی $A = P \text{diag}(\lambda_1, \dots, \lambda_m) P^T$.

ب. برای هر $(A, B) \in S_+^m \times S_+^m$ داریم $\langle A, B \rangle = 0$ اگر و فقط اگر $A = [A - B]_+$.

لم ۳.۲.۲. اگر $A, B \in S^m$ اگر $A \succeq 0$ و $B \succeq 0$ آنگاه $\langle A, B \rangle = 0$ اگر و فقط اگر $AB = 0$.

۳.۲ برنامه ریزی نیمه معین

مساله بهینه سازی زیر را در نظر بگیرید:

$$\text{minimize } f(x) = c^T x \quad (۴.۲)$$

$$\text{subject to } A(x) = A_0 + \sum_{i=1}^m x_i A_i \preceq 0, \quad (۵.۲)$$

که $c \in \mathbb{R}^n$ ، $x \in \mathbb{R}^n$ ، $A_i \in \mathbb{R}^{m \times m}$ ، $(i = 1, 2, \dots, n)$. نامساوی $A(x) \preceq 0$ را یک نامساوی ماتریسی خطی و مساله (۴.۲) و (۵.۲) را یک مساله نیمه معین خطی می نامند.

برنامه ریزی نیمه معین یک مساله بهینه سازی محدب است زیرا تابع هدف و محدودیت

آن محدب هستند، یعنی اگر $A(x) \preceq 0$ و $A(y) \preceq 0$ آنگاه به ازای هر λ که $0 < \lambda < 1$ داریم

$$A(\lambda x + (1 - \lambda)y) = \lambda A(x) + (1 - \lambda)A(y) \preceq 0.$$

اگرچه برنامه ریزی نیمه معین (۴.۲) و (۵.۲) ممکن است کاملاً ویژه ظاهر شده باشد، ما خواهیم دید که آن شامل بسیاری از مسائل بهینه سازی مهم به عنوان حالت خاص است. برای نمونه برنامه ریزی خطی را در نظر بگیرید

$$\text{minimize } f(x) = c^T x \quad (۶.۲)$$

$$\text{subject to } Ax + b \geq 0. \quad (۷.۲)$$

چون یک بردار $v \geq \circ$ اگر و تنها اگر ماتریس $diag(v)$ (ماتریس قطری با مولفه‌های v روی قطرش) نیمه معین مثبت باشد، ما می‌توانیم برنامه‌ریزی خطی (۶.۲)–(۷.۲) را به عنوان یک برنامه‌ریزی نیمه معین با $A(x) = diag(Ax + b)$ بیان کنیم. یعنی

$$A = [a_1 \dots a_n] \in \mathbb{R}^{m \times n} \text{ که } (i = 1, \dots, n), A_i = diag(a_i), A_0 = diag(b)$$

برنامه‌ریزی نیمه معین می‌تواند به عنوان یک توسیع از برنامه‌ریزی خطی دیده شود که نامساوی مولفه به مولفه بین بردارها با نامساوی‌های ماتریسی عوض می‌شود. همچنین می‌توانیم برنامه ریزی نیمه معین را به عنوان یک برنامه ریزی نیمه نامتناهی مشاهده کنیم، زیرا نامساوی ماتریسی $A(x) \leq \circ$ با یک مجموعه نامتناهی از محدودیت‌های خطی روی x معادل است، یعنی $\circ z^T A(x) z \leq \circ$ به ازای هر $z \in \mathbb{R}^n$. بنابراین تعجبی ندارد که نظریه برنامه ریزی نیمه معین به طور نزدیکی موازی با برنامه ریزی خطی است، یا این که بسیاری از الگوریتم‌ها برای حل برنامه‌ریزی خطی باید تعمیم داده شوند تا برنامه‌ریزی نیمه معین را حل کنند. به‌علاوه، چند تفاوت مهم وجود دارد: نتایج دوگانی برای برنامه‌ریزی‌های نیمه معین ضعیف‌تر از نتایج دوگانی برای برنامه‌ریزی‌های خطی هستند و هیچ روش مستقیم یا روش سیمپلکس عملی برای حل برنامه‌ریزی‌های نیمه معین وجود ندارد.

قبل از آنکه جلوتر برویم یک مثال ساده از یک مساله بهینه‌سازی غیرخطی (محدب) که می‌تواند به عنوان یک برنامه‌ریزی نیمه معین طرح شود، اما به عنوان یک برنامه‌ریزی خطی نیست، ارائه می‌دهیم. مساله زیر را در نظر بگیرید

$$\text{minimize } \frac{(c^T x)^2}{d^T x} \quad (۸.۲)$$

$$\text{subject to } Ax + b \geq \circ. \quad (۹.۲)$$

فرض می‌کنیم که $d^T x > \circ$ وقتی که $Ax + b \geq \circ$. با در نظر گرفتن یک متغیر t مساله بالا به مساله معادل زیر تبدیل می‌شود:

$$\text{minimize } t \quad (۱۰.۲)$$

$$\text{subject to } Ax + b \geq \circ. \quad (۱۱.۲)$$

$$\frac{(c^T x)^2}{d^T x} \leq t. \quad (۱۲.۲)$$

تابع هدف غیرخطی (محدب) در (۸.۲) به عنوان یک محدودیت غیرخطی (محدب) در (۱۲.۲) نشان داده شده است. به این ترتیب این محدودیت‌ها می‌توانند به عنوان یک نامساوی ماتریسی خطی در متغیرهای x و t بیان شوند

$$\text{minimize } t \quad (۱۳.۲)$$

$$\text{subject to } \begin{pmatrix} diag(Ax + b) & \circ & \circ \\ \circ & t & c^T x \\ \circ & c^T x & d^T x \end{pmatrix} \succeq \circ. \quad (۱۴.۲)$$

پس مساله برنامه ریزی غیرخطی (محدب) (۸.۲) و (۹.۲) را به عنوان یک برنامه ریزی نیمه معین (۱۳.۲) و (۱۴.۲) فرمول بندی کرده ایم. این نامساوی ماتریسی خطی در (۱۴.۲) دو نکته را نشان می دهد

۱. نمایش چند نامساوی ماتریسی خطی به عنوان یک نامساوی ماتریسی قطری بلوکی

۲. استفاده از (۱.۲) برای نمایش یک محدودیت محدب غیرخطی به عنوان یک نامساوی ماتریسی خطی

توجه کنید که با استفاده از (۱.۲) نامساوی ماتریس 2×2

$$\begin{pmatrix} t & c^T x \\ c^T x & d^T x \end{pmatrix} \succeq 0. \quad (15.2)$$

معادل است با $d^T x \geq 0$ و $t - \frac{(c^T x)^2}{d^T x} \geq 0$ (با $t \geq 0$ ، اگر $c^T x = 0$ ، $d^T x = 0$). چون فرض کرده ایم که $Ax + b \geq 0$ نتیجه می دهد $d^T x > 0$. می بینیم که محدودیت (۱۱.۲) و (۱۲.۲) با نامساوی ماتریسی (۱۴.۲) معادل هستند. عبارت $t - \frac{(c^T x)^2}{d^T x}$ مکمل شور $d^T x$ در نامساوی ماتریسی (۱۵.۲) نامیده می شود. تشخیص مکمل های شور نکته ای کلیدی در فرمول بندی دوباره مسائل بهینه سازی محدب غیرخطی به عنوان برنامه ریزی نیمه معین است.

دلایل خوبی برای مطالعه برنامه ریزی نیمه معین وجود دارد. اول محدودیت های نیمه معین به طور مستقیم در کاربردهای مهم بسیاری نتیجه می شوند. دوم این که بسیاری از مسائل بهینه سازی محدب برای مثال برنامه ریزی خطی و برنامه ریزی درجه دوم به صورت درجه دوم مقید شده (محدب) می توانند به عنوان برنامه ریزی نیمه معین طرح شوند. در نتیجه برنامه ریزی نیمه معین یک راه یکنواخت شده برای مطالعه خواص و الگوریتم های بدست آمده برای یک حوزه گسترده از مسائل بهینه سازی محدب پیشنهاد می کند. علاوه بر آن، مهم تر اینکه برنامه ریزی نیمه معین می تواند خیلی مؤثر، هم در نظریه و هم در عمل حل شود. نظریه آسانی از محدب بودن نتیجه می شود.

۴.۲ مثال ها

در این بخش تعدادی مثال و کاربرد برنامه ریزی نیمه معین را بیان می کنیم. لیست کامل نیست و طرز عمل ما از کاربردهای برنامه ریزی نیمه معین شتابزده است. بازنگری برنامه ریزی نیمه معین در حوزه های گوناگون تاکنون منتشر شده است. برای نظریه کنترل [۱۴]، بهینه سازی ترکیبیاتی [۳]، آمار [۷۷] و بهینه سازی ساختاری [۱۱] را نگاه کنید. هدف ما ارائه یک ایده کلی از مسائل برای خاطر نشان کردن ارتباط بین کاربردها در حوزه های گوناگون و فراهم کردن مرجعی برای مطالعات آینده است. همچنین نسترو و نمیروفسکی و بوید [۷۲، ۱۴] را برای مثال های بیشتر نگاه کنید.

۱.۴.۲ برنامه‌ریزی درجه دوم به صورت درجه دوم مقید

یک محدودیت درجه دوم محدب

$$(Ax + b)^T(Ax + b) - c^T x - d \leq 0$$

با $x \in \mathbb{R}^k$ طبق لم شور (۱.۲) می‌تواند به صورت زیر نوشته شود

$$\begin{pmatrix} I & Ax + b \\ (Ax + b)^T & c^T x + d \end{pmatrix} \succeq 0.$$

سمت چپ به طور آفینی به بردار x بستگی دارد، که می‌تواند به صورت زیر بیان شود

$$F(x) = F_0 + x_1 F_1 + \dots + x_k F_k \succeq 0,$$

با

$$F_0 = \begin{pmatrix} I & b \\ b^T & d \end{pmatrix}, \quad F_i = \begin{pmatrix} 0 & a_i \\ a_i^T & c_i \end{pmatrix}, \quad i = 1, \dots, k,$$

که در آن $A = [a_1 \dots a_k]$. بنابراین یک برنامه‌ریزی درجه دوم به صورت مقید^{۱۵} (محدب) کلی

$$\begin{aligned} & \text{minimize } f_0(x) \\ & \text{subject to } f_i \leq 0, \quad i = 1, \dots, L, \end{aligned}$$

که هر f_k یک تابع درجه دوم محدب

$$f_k(x) = (A_k x + b)^T(A_k x + b) - c_k^T x - d_k, \quad k = 0, 1, \dots, L$$

می‌تواند به صورت زیر نوشته شود

$$\begin{aligned} & \text{minimize } t \\ & \text{subject to } \begin{pmatrix} I & A_0 x + b_0 \\ (A_0 x + b_0)^T & c_0^T + d_0 + t \end{pmatrix} \succeq 0, \\ & \begin{pmatrix} I & A_i x + b_i \\ (A_i x + b_i)^T & c_i^T + d_i \end{pmatrix} \succeq 0, \quad i = 1, \dots, L, \end{aligned}$$

که یک برنامه‌ریزی نیمه معین با متغیرهای $x \in \mathbb{R}^k$ و $t \in \mathbb{R}$ است. این برنامه‌ریزی نیمه معین ابعاد $m = k + 1$ و $n = n_0 + \dots + n_L$ را دارد که $A_i \in \mathbb{R}^{n_i}$.

^{۱۵}Quadratically Constrained Quadratic Programming(QCQP)

۲.۴.۲ بزرگترین مقدار ویژه و مینیمم سازی نرم ماتریسی

فرض کنید ماتریس متقارن $A(x)$ به طور آفینی روی $x \in \mathbb{R}^k$ تعریف شده است، یعنی $A(x) = A_0 + x_1 A_1 + \dots + x_k A_k$ که در آن $A_i = A_i^T \in \mathbb{R}^{p \times p}$ مساله مینیمم سازی مسئله بزرگترین مقدار ویژه ماتریس $A(x)$ می تواند به عنوان برنامه ریزی نیمه معین به صورت زیر طرح شود:

$$\begin{aligned} & \text{minimize } \lambda \\ & \text{subject to } \lambda I - A(x) \succeq 0, \end{aligned}$$

با متغیرهای $(x, \lambda) \in \mathbb{R}^{k+1}$. مساله فوق کوچکترین λ را می یابد به طوری که λ از همه مقادیر ویژه ی ماتریس $A(x)$ بزرگتر است. اما λ بزرگترین مقدار ویژه است اگر و فقط اگر $\lambda I - A(x)$ نیمه معین مثبت باشد. مسائلی از این نوع در نظریه کنترل، نظریه گراف، بهینه سازی ساختاری، بهینه سازی ترکیبیاتی، آمار و سایر حوزه ها ظاهر می شوند. اورتون^{۱۶}، مهار^{۱۷} و پلجک^{۱۸} [۷۲، ۸۸] را ببینید.

چندین مساله جالب وابسته می تواند با استفاده از برنامه ریزی نیمه معین حل شوند. به عنوان مثال، برای مینیمم کردن مجموع r تا بزرگترین مقادیر ویژه $A(x)$ ، می توان برنامه ریزی نیمه معین زیر را حل کرد

$$\begin{aligned} & \text{minimize } rt + Tr(X) \\ & \text{subject to } \lambda t I + X - A(x) \succeq 0, \\ & X \succeq 0, \end{aligned}$$

که در آن $Tr(X) = X_{11} + \dots + X_{pp}$ تریس (اثر) ماتریس $X \in \mathbb{R}^p$ را نشان می دهد. برای اثبات علیزاده [۵] را ببینید.

همچنین مسئله مینیمم کردن فاصله بین بزرگترین و کوچکترین مقدار ویژه ماتریس $A(x)$ می تواند به عنوان برنامه ریزی نیمه معین زیر بیان شود

$$\begin{aligned} & \text{minimize } \lambda - \mu \\ & \text{subject to } \mu I \preceq A(x) \preceq \lambda I, \end{aligned}$$

که $\mu = \lambda_{\min}(A(x))$ و $\lambda = \lambda_{\max}(A(x))$.

این نتایج همچنین می تواند به مسائل شامل قدر مطلق مقادیر ویژه یا مجموع وزن دار از مقادیر ویژه تعمیم داده شود. علیزاده [۵] را ببینید.

^{۱۶}Overton

^{۱۷}Mohar

^{۱۸}Poljak

مساله وابسته دیگر مینیمم کردن نرم ماتریس $\|A(x)\|$ (نرم طیفی^{۱۹} یا ماکزیمم مقدار تکین) از ماتریس $A(x) = A_0 + x_1 A_1 + \dots + x_k A_k \in \mathbb{R}^p$ است (اینجا نیاز نیست A_i ها متقارن باشند). یادآوری می‌کنیم که نرم طیفی یک ماتریس به صورت $\|A\| = \sqrt{\lambda_{\max}(A^T A)}$ تعریف می‌شود. این مساله می‌تواند به عنوان برنامه‌ریزی نیمه معین به صورت زیر طرح شود

$$\begin{aligned} & \text{minimize } t \\ & \text{subject to } \begin{pmatrix} tI & A(x) \\ A(x)^T & tI \end{pmatrix} \succeq 0, \end{aligned}$$

با متغیرهای $x \in \mathbb{R}^k$ و $t \in \mathbb{R}$.

توجه کنید که توابع هدف در این مسائل یعنی بزرگترین مقدار ویژه، مجموع r تا بزرگترین مقادیر ویژه و نرم، توابع مشتق‌ناپذیر از x (اما محدب) هستند.

۳.۴.۲ تقریب لگاریتمی چبیشف

فرض کنید که علاقمندیم که مسئله $A(x) \approx b$ را به طور تقریبی حل کنیم، که در آن $A = [a_1, \dots, a_p] \in \mathbb{R}^{p \times k}$ و $b \in \mathbb{R}^p$. در تقریب چبیشف ما نرم l_∞ باقیمانده را مینیمم می‌کنیم. یعنی

$$\min \max_i |a_i^T x - b_i|.$$

این مساله می‌تواند به عنوان یک برنامه‌ریزی خطی با متغیرهای x و t به عنوان متغیرهای مصنوعی به صورت زیر طرح شود

$$\begin{aligned} & \text{minimize } t \\ & \text{subject to } -t \leq a_i^T x - b_i \leq t, \quad i = 1, \dots, p. \end{aligned}$$

در چند کاربرد b_i بعدی از توان یا شدت دارد و به صورت یک مقیاس لگاریتمی بیان می‌شود. در چنین حالت‌هایی مساله بهینه‌سازی طبیعی به صورت زیر است

$$\min \max_i | \log(a_i^T x) - \log(b_i) |. \quad (۱۶.۲)$$

(فرض کنید $b_i > 0$ و $\log(a_i^T x)$ به صورت $-\infty$ است وقتی که $a_i^T x \leq 0$). این مساله تقریب لگاریتمی چبیشف^{۲۰} می‌تواند به صورت یک برنامه‌ریزی نیمه معین طرح شود. برای دیدن این موضوع توجه کنید که

$$| \log(a_i^T x) - \log(b_i) | = \log \max \left(\frac{a_i^T x}{b_i}, \frac{b_i}{a_i^T x} \right).$$

^{۱۹} Spectral norm

^{۲۰} Logarithmic Chebychev Approximation

(فرض کنید که $a_i^T x > 0$). بنابراین مساله (۱۶.۲) معادل مساله زیر است

$$\begin{aligned} & \text{minimize } t \\ & \text{subject to } \frac{1}{t} \leq \frac{a_i^T x}{b_i} \leq t, \quad i = 1, \dots, p. \end{aligned}$$

یا

$$\begin{aligned} & \text{minimize } t \\ & \text{subject to } \begin{pmatrix} t - \frac{a_i^T x}{b_i} & 0 & 0 \\ 0 & \frac{a_i^T x}{b_i} & 1 \\ 0 & 1 & t \end{pmatrix} \succeq 0, \quad i = 1, \dots, p. \end{aligned}$$

که یک برنامه ریزی نیمه معین است. این مثال دو نکته مهم را نشان می دهد. اول برنامه ریزی نیمه معین شامل بسیاری از مسائل بهینه سازی است که در اولین نگاه مانند مساله نیمه معین به نظر نمی رسند و دوم برنامه ریزی نیمه معین از برنامه ریزی خطی کلی تر است، علیرغم شباهت نزدیکی که به هم دارند.

۴.۴.۲ بهینه سازی ساختاری

بن-تال^{۲۱} و بندسی^{۲۲} در [۱۱] مساله بهینه سازی ساختاری^{۲۳} زیر را در نظر گرفته اند. یک ساختار از k میله کشسان یک مجموعه از p گره را به هم متصل می کند. هندسه (مکان شناسی و طول میله ها) و مواد ثابت هستند. هدف اندازه میله ها است. یعنی معین کردن تقریبی نواحی پیوندی میله ها. در ساده ترین نسخه از مساله ما یک مجموعه ثابت از نیروهای گره های به طور خارجی به کار برده شده f_i ، $(i = 1, \dots, p)$ را در نظر می گیریم. بردار تغییر مکان های گره ها که از حمل نیروهای f_i نتیجه می شود، با d نشان داده می شود. هدف این است که انرژی ذخیره شده کشسان $\frac{1}{2} f^T d$ مینیمم گردد. ما همچنین نیاز داریم که محدودیت های را روی اندازه مجموع (یا به طور معادل وزن) و کران های بالا و پایین روی ناحیه پیوندی روی هر میله به حساب آوریم.

متغیرهای طراحی ناحیه های پیوندی x_i هستند. وابستگی بین f و d خطی است، یعنی

$$f = A(x)d \quad A(x) = \sum_{i=1}^k x_i A_i$$

ماتریس سختی^{۲۴} نامیده می شود. ماتریس های A_i همگی نیمه معین مثبت متقارن هستند و تنها به پارامترهای ثابت بستگی دارند (طول میله ها و هندسه). پس مساله بهینه سازی به

^{۲۱} Ben-Tal

^{۲۲} Bendse

^{۲۳} Structural optimization

^{۲۴} Stiffness matrix

صورت زیر است

$$\begin{aligned} & \text{minimize } f^T d \\ & \text{subject to } f = A(x)d, \\ & \quad \sum_{i=1}^k l_i x_i \leq \nu, \\ & \quad \underline{x}_i \leq x_i \leq \bar{x}_i, i = 1, \dots, k, \end{aligned}$$

که در آن d و x متغیرها هستند، ν بزرگترین اندازه است، l_i طول میله‌ها و $\underline{x}_i, \bar{x}_i$ کران‌های پایین و بالا روی ناحیه پیوندی هستند. برای ساده‌سازی ما فرض می‌کنیم که $\underline{x}_i > 0$ و $A(x) > 0$ به ازای همه مقادیر مثبت x_i . پس ما می‌توانیم d را تخمین بزنیم و بنویسیم

$$\begin{aligned} & \text{minimize } f^T A(x)^{-1} f \\ & \text{subject to } \sum_{i=1}^k l_i x_i \leq \nu, \\ & \quad \underline{x}_i \leq x_i \leq \bar{x}_i, i = 1, \dots, k, \end{aligned}$$

یا

$$\begin{aligned} & \text{minimize } t \\ & \text{subject to } \begin{pmatrix} t & f^T \\ f & A(x) \end{pmatrix} \succeq 0, \quad i = 1, \dots, p, \\ & \quad \sum_{i=1}^k l_i x_i \leq \nu, \\ & \quad \underline{x}_i \leq x_i \leq \bar{x}_i, i = 1, \dots, k, \end{aligned}$$

که یک برنامه‌ریزی نیمه معین در x و t است (توجه کنید که از لم شور (۱.۲) برای بیان $f^T A(x)^{-1} f \leq 0$ به عنوان یک نامساوی ماتریسی خطی استفاده کرده‌ایم).

۵.۴.۲ کاربرد در آمار

فرض کنید که $x \in \mathbb{R}^p$ یک بردار تصادفی با میانگین \bar{x} و ماتریس کوواریانس Σ باشد. تعداد زیادی از نمونه‌های $y = x + n$ را می‌گیریم که نویز اندازه‌گیری n ، میانگین صفر دارد، به x ارتباطی ندارد و یک ماتریس کوواریانس نامعلوم اما قطری D دارد. در نتیجه $\hat{\Sigma} = \Sigma + D$ که ماتریس کوواریانس y را نشان می‌دهد. فرض می‌کنیم که می‌توانیم Σ را با اطمینان بالا تخمین بزنیم، یعنی آن را یک ماتریس معلوم و ثابت در نظر می‌گیریم. ماتریس Σ ، ماتریس کوواریانس x ، یا ماتریس D ، ماتریس کوواریانس نویز اندازه‌گیری را نمی‌شناسیم. اما آن‌ها هر دو نیمه معین مثبت هستند بنابراین می‌دانیم که Σ در مجموعه محدب زیر قرار دارد

$$\Sigma = \{ \hat{\Sigma} - D \mid \hat{\Sigma} - D \succeq 0, D > 0, D \text{ diagonal} \}$$

این حقیقت به ما اجازه می دهد که کرانها را روی توابع خطی Σ با حل مسائل برنامه ریزی نیمه معین روی مجموعه Σ به دست بیاوریم. به عنوان یک مثال، مجموع مؤلفه های x یعنی $e^T x$ که برداری با تمام مؤلفه های برابر یک است، در نظر بگیرید. واریانس $e^T x$ با

$$e^T \hat{\Sigma} e = e^T \Sigma e - Tr(D)$$

در نظر گرفته می شود. با توجه به این که ماتریس Σ نامعلوم است پس $e^T \hat{\Sigma} e$ نیز نامعلوم است. اما با تعیین ماکزیمم و مینیمم $e^T \hat{\Sigma} e$ روی مجموعه Σ می توانیم یک بازه را معین کنیم که در آن قرار دارد. به عبارت دیگر، ما می توانیم کران پایین و بالا را روی $e^T \hat{\Sigma} e$ محاسبه کنیم. به آسانی دیده می شود که کران بالا $e^T \hat{\Sigma} e$ است. یک کران پایین می تواند با حل برنامه ریزی نیمه معین زیر محاسبه شود

$$\text{minimize } \sum_{i=1}^p d_i \quad (17.2)$$

$$\text{subject to } \Sigma - \text{diag}(d) \succeq 0, \quad (18.2)$$

$$d \geq 0. \quad (19.2)$$

فلچر [۲۹] (۱۷.۲)–(۱۹.۲) را مساله آزمون آزمایشی^{۲۵} نامید. بردار y نمرات یک دانش آموز که به طور تصادفی انتخاب می شود روی یک سری از p آزمون را نشان می دهد و $e^T y$ مجموع نمرات را نشان می دهد. شخص آزمون را معتبر در نظر می گیرد اگر واریانس مجموع نمرات اندازه گیری شده $e^T y$ نزدیک واریانس $e^T x$ روی کل جمعیت باشد. مقدار

$$\rho = \frac{e^T \Sigma e}{e^T \hat{\Sigma} e}$$

اعتبار آزمون^{۲۶} نامیده می شود. با حل برنامه ریزی نیمه معین (۱۷.۲)–(۱۹.۲) می توان یک کران پایین برای ρ محاسبه کرد.

برنامه ریزی نیمه معین همچنین کاربردهایی در طراحی آزمایش بهینه دارد ([۷۷] را ببینید). علاوه بر مثال هایی که ارائه شد، مسائل برنامه ریزی نیمه معین به طور مکرر در نظریه کنترل و سیستم به وجود می آیند. بوید، ال قایی^{۲۷}، فرون^{۲۸} و بالا کریشن^{۲۹} مثال های بسیاری در [۸۸] جمع آوری کرده اند. همچنین بسیاری از مسائل هندسی شامل فرم های درجه دوم و مسائل بهینه سازی ترکیبیاتی^{۳۰} و نامحدب نیز می توانند به عنوان مسئله برنامه ریزی نیمه معین طرح شوند ([۸۸] را ببینید).

^{۲۵}Educational testing problem

^{۲۶}Reliability of the test

^{۲۷}El Ghaoui

^{۲۸}Fron

^{۲۹}Balakrishnan

^{۳۰}Combinatorial optimization

۵.۲ دوگانی در برنامه‌ریزی نیمه معین

دافین^{۳۱} در [۱۹] اولین کسی بود که قضایای دوگان تعمیم یافته را برای مسئله نیمه معین مطالعه کرد. پس از او هارویکز^{۳۲} [۴۱] باروین^{۳۳} و ولکوویکز^{۳۴} [۱۲] و ولکوویکز [۹۲] فرمول‌بندی قضیه دوگان را گسترش دادند. اندرسن^{۳۵} و نش^{۳۶} در [۶] قضیه دوگان تعمیم یافته را با دید مساله خطی با بعد نامتناهی ارائه دادند. علیزاده در [۵] قضایای دوگانی مسئله نیمه معین را بیان و اثبات نمود. برنامه‌ریزی نیمه معین (۴.۲) و (۵.۲) را به عنوان مساله اولیه در نظر بگیرید. دوگان متناظر با آن به صورت زیر تعریف می‌شود

$$\text{maximize } \langle A_0, Z \rangle \quad (20.2)$$

$$\text{subject to } \langle -A_i, Z \rangle = c_i, \quad i = 1, \dots, n, \quad (21.2)$$

$$Z \succeq 0, \quad (22.2)$$

که در آن ماتریس $Z \in S^m$. توجه کنید که (۲۰.۲) یک تابع هدف خطی از Z است. می‌گوییم دوگان شدنی است اگر $TrZA_i = -c_i$ ، $i = 1, \dots, n$ و $Z \succeq 0$. خاصیت کلیدی برنامه‌ریزی نیمه معین دوگان این است که کران‌هایی روی مقدار بهینه از برنامه‌ریزی نیمه معین اولیه نتیجه می‌دهد و برعکس. فرض کنید که Z دوگان شدنی و x اولیه شدنی باشد، پس

$$c^T x - Tr(ZA_0) = -\sum_{i=1}^n Tr(ZA_i x_i) - Tr(ZA_0) = -Tr(ZA(x)) \geq 0, \quad (23.2)$$

که از $Tr(AB) \geq 0$ وقتی که $A = A^T \succeq 0$ و $B = B^T \succeq 0$ استفاده می‌کنیم. بنابراین داریم

$$c^T x \geq Tr(ZA_0), \quad (24.2)$$

یعنی مقدار تابع هدف هر نقطه شدنی دوگان Z کوچکتر یا مساوی با مقدار تابع هدف اولیه هر نقطه شدنی اولیه x است. تفاوت را به عنوان فاصله دوگانی η متناظر با x و Z به صورت زیر در نظر می‌گیریم

$$\eta = c^T x - Tr(ZA_0) = Tr(ZA(x)).$$

توجه کنید که فاصله دوگانی یک تابع خطی از x و Z است. فرض کنید p^* مقدار بهینه برنامه‌ریزی نیمه معین (۴.۲) و (۵.۲) را نشان دهد، یعنی

$$p^* = \inf \{c^T x | A(x) \preceq 0\}$$

^{۳۱}Duffin

^{۳۲}Hurwicz

^{۳۳}Borwein

^{۳۴}Wolkowicz

^{۳۵}Anderson

^{۳۶}Nash

و فرض کنید دوگان شدنی باشد. چون (۲۴.۲) برای هر x شدنی برقرار است، نتیجه می‌گیریم که $p^* \geq Tr(ZA_0)$. به عبارت دیگر ماتریس‌های دوگان کران‌های پایین را برای مساله اولیه نتیجه می‌دهند.

به طور مشابه، نقاط شدنی اولیه کران‌های بالا برای مساله دوگان را نتیجه می‌دهند یعنی $d^* \leq c^T x$ که مقدار بهینه مسئله دوگان نیمه معین (۲۰.۲)–(۲۲.۲) است و به صورت زیر به دست می‌آید

$$d^* = \sup\{Tr(ZA_0) | Z \succeq 0, -Tr(ZA_i) = c_i, \quad i = 1, \dots, n\}$$

نتیجه می‌شود که $d^* \leq p^*$ ، یعنی مقدار بهینه مساله دوگان کمتر یا مساوی با مقدار بهینه مساله اولیه است. در واقع تساوی معمولاً بدست می‌آید اما ممکن است در برخی مسائل فاصله دوگانی در بهینگی صفر نباشد. فرض کنید X_{opt} و Z_{opt} مجموعه‌های بهینه اولیه و دوگان را نشان می‌دهند، یعنی

$$X_{opt} = \{x | A(x) \preceq 0, \quad c^T x = p^*\},$$

$$Z_{opt} = \{Z | Z \succeq 0, -Tr(ZA_i) = c_i, \quad i = 1, \dots, n, \quad Tr(ZA_0) = d^*\}$$

توجه کنید که X_{opt} (یا Z_{opt}) می‌تواند تهی باشد، حتی اگر p^* (یا d^*) متناهی باشد، مانند برنامه‌ریزی نیمه معین زیر

$$\begin{aligned} & \text{minimize } t \\ & \text{subject to } \begin{pmatrix} x & 1 \\ 1 & t \end{pmatrix} \succeq 0. \end{aligned}$$

قضیه ۱.۵.۲. [۷۲]، [۸۰] $p^* = d^*$ اگر هر یک از شرایط زیر برقرار باشند

۱. مساله اولیه (۴.۲) و (۵.۲) شدنی باشد، یعنی یک x با $A(x) \prec 0$ وجود داشته باشد.
۲. مساله دوگان (۲۰.۲)–(۲۲.۲) شدنی باشد، یعنی یک Z با $Z = Z^T \succ 0$ ، $Tr A_i Z = -c_i$ ، $i = 1, \dots, n$ وجود داشته باشد.

اگر هر دو شرط برقرار باشند، مجموعه‌های بهینه X_{opt} و Z_{opt} ناتهی هستند.

ولکوویز [۹۲] و رامانا^{۳۷} [۷۹] دو رهیافت مختلف برای نظریه دوگانی برای برنامه‌ریزی نیمه معین مطرح کرده‌اند که نیاز نیست اکیدا شدنی باشد. برای اهداف کنونی ما دوگانی استاندارد نمادگذاری شده در بالا کافی خواهد بود.

شرایط بهینگی KKT

فرض کنید که مجموعه‌های بهینه اولیه و دوگان ناتهی باشند، یعنی x و Z با

$$p^* = c^T x = TrZA_0 = d^*$$

وجود داشته باشند. از (۲۳.۲) داریم $-TrZA(x) = 0$. چون $-A(x) \succeq 0$ و $Z \succeq 0$ از لم (۳.۲.۲) نتیجه می‌گیریم که $ZA(x) = 0$. شرط $TrZA(x) = 0$ شرط مکمل زائد است و بیان می‌کند که فضای برد ماتریس‌های متقارن Z و $A(x)$ متعامد هستند. این شرط مکمل زائد برنامه‌ریزی خطی را عمومیت می‌بخشد. قضیه (۱.۵.۲) شرایط بهینگی برای برنامه‌ریزی نیمه معین (۴.۲) و (۵.۲) را نتیجه می‌دهد. شرایط لازم بهینگی KKT برای یک نقطه شدنی x_* از مسئله نیمه معین (۴.۲) و (۵.۲) وجود یک $Z_* \in S^m$ است به طوری که

$$\nabla f(x_*) + DA(x_*)^* Z_* = 0, \quad (25.2)$$

$$\langle Z_*, A(x_*) \rangle = 0, \quad (26.2)$$

$$A(x_*) \preceq 0, \quad Z_* \succeq 0. \quad (27.2)$$

x_* نقطه KKT مسئله نیمه معین (۴.۲) و (۵.۲) و Z_* ضربگر لاگرانژ متناظر با x_* نامیده می‌شود.

یک مثال ساده با $p^* \neq d^*$ را به صورت زیر در نظر می‌گیریم

$$\begin{aligned} & \text{minimize } x_1 \\ & \text{subject to } \begin{pmatrix} 0 & x_1 & 0 \\ x_1 & x_2 & 0 \\ 0 & 0 & x_1 + 1 \end{pmatrix} \succeq 0. \end{aligned}$$

مجموعه شدنی $\{(x_1, x_2)^T | x_1 = 0, x_2 \geq 0\}$ است، بنابراین $p^* = 0$. برنامه‌ریزی دوگان به صورت زیر بیان می‌شود

$$\begin{aligned} & \text{maximize } -z_2 \\ & \text{subject to } \begin{pmatrix} z_1 & \frac{(1-z_2)}{2} & 0 \\ \frac{(1-z_2)}{2} & 0 & 0 \\ 0 & 0 & z_2 \end{pmatrix} \succeq 0. \end{aligned}$$

که مجموعه شدنی $\{(z_1, z_2)^T | z_2 = 1, z_1 \geq 0\}$ را دارد. پس مساله دوگان جواب بهینه $d^* = -1$ را دارد. البته، این برنامه‌ریزی نیمه معین از هر دو شرط در قضیه (۱.۵.۲) تخطی می‌کند. هر دو مساله شدنی هستند، اما اکیدا شدنی نیستند. همچنین توجه کنید که برخلاف

برنامه ریزی خطی می تواند یک فاصله دوگانی متناهی غیر صفر در بهینگی وجود داشته باشد. بنابراین در تمام این رساله فرض می کنیم دو شرط زیر برقرار هستند

۱. ماتریس های A_i ، $i = 1, \dots, n$ ، مستقل خطی هستند، یعنی $\sum_{i=1}^n u_i A_i = 0$ نتیجه می دهد $u_i = 0$ ، $i = 1, \dots, n$.

۲. هر دو مسئله اولیه و دوگان شدنی هستند، یعنی

$$\exists Z_0 \in S^m \quad \text{and} \quad x_0 \in \mathbb{R}^n : Z_0 \succ 0, \quad A(x_0) \prec 0, \quad \langle -A_i, Z_0 \rangle = c_i, i = 1, \dots, n.$$

فصل ۳

یک مدل شبکه عصبی برای حل مسئله بهینه سازی نیمه معین خطی

۱.۳ مقدمه

مسئله نیمه معین در سال‌های اخیر بسیار مورد توجه قرار گرفته و کاربردهای گسترده‌ای در مهندسی و بهینه‌سازی ساختاری^۱ یافته است. هر چند در نگاه اول ممکن است مسئله نیمه معین خاص به نظر برسد، بسیاری از مسائل بهینه‌سازی به عنوان حالت خاصی از مسائل نیمه معین هستند، مانند برنامه‌ریزی خطی، مینیمم کردن بزرگترین مقدار ویژه ماتریس، بیشترین دترمینان و تقریب لگاریتمی چیشف ([۸۸] را ببینید). این مسائل در نظریه کنترل، نظریه گراف، آمار، بهینه‌سازی ساختاری، بهینه‌سازی ترکیبیاتی و سایر زمینه‌ها ظاهر می‌شوند. ([۵، ۶۷، ۷۴، ۸۸] را ببینید).

مسئله نیمه معین تعمیم مسئله برنامه‌ریزی خطی است که در آن نامساوی بین بردارها با نامساوی بین ماتریس‌ها جایگزین شده است. به همین دلیل بسیاری از روش‌های نقطه درونی که برای حل مسائل برنامه‌ریزی خطی طراحی شده‌اند، به طور موفق‌تری برای حل مسئله برنامه‌ریزی نیمه معین به کار رفته‌اند ([۳، ۳۵، ۵۱، ۵۴، ۸۷، ۹۰، ۱۰۸] را ببینید). در واقع، ثابت شده است که روش‌های نقطه درونی برای حل مسائل با بعد متوسط بسیار قابل استفاده‌اند.

^۱Structural optimization

اما اگر مساله روی ماتریس‌هایی با بعد بزرگ تعریف شود یا تعداد محدودیت‌ها زیاد باشد، رشد روش‌های نقطه درونی بسیار کند بوده و برای حل آن نیاز به حافظه‌ی زیادی است. در [۳۵] و [۹۱] روش‌های جهت متناوب لاگرانژی^۲ و دسته‌بندی طیفی^۳ برای مساله نیمه معین ارائه شده است. با این حال، جواب‌های زمان واقعی در بسیاری از مسائل کاربردی مانند تحلیل نیرو در برنامه ریزگری^۴ و کنترل ربات‌ها [۴۴، ۱۱۱] ضروری است. روش‌های سنتی بهینه‌سازی ممکن است برای این مسائل مناسب نباشد زیرا زمان دقیق محاسبات مورد نیاز است. بنابراین، برای حل مسائل بهینه‌سازی در زمان واقعی، یک روش قابل قبول و کارآمد نیاز است.

روش شبکه عصبی یک روش ایده‌آل برای حل مسائل بهینه‌سازی در زمان واقعی است. در مقایسه با روش‌های قبلی، روش شبکه عصبی در حل مسائل بهینه‌سازی در زمان واقعی، دارای مزایایی است. اول ساختار یک شبکه عصبی می‌تواند به طور موثر با استفاده از یکپارچه‌سازی مقیاس بزرگ^۵ و تکنولوژی‌های نوری^۶ به کار برده شود. دوم، شبکه عصبی می‌تواند بسیاری از مسائل بهینه‌سازی را با پارامترهای متغیر با زمان حل کند. سوم، روش‌های دینامیکی و معادلات دیفرانسیل عددی معمولی^۷ می‌توانند به طور موثر و مستقیم برای حل مسائل بهینه‌سازی مقید به وسیله‌ی شبکه عصبی زمان پیوسته به کار گرفته شوند. چهارم، شبکه‌های عصبی سرعت همگرایی بالایی در جواب‌های زمان واقعی دارند. به همین دلیل روش‌های شبکه عصبی برای بهینه‌سازی مورد توجه قرار گرفته‌اند. از این رو، محققان بسیاری شبکه‌های عصبی پیوسته را برای مسائل بهینه‌سازی مقید ایجاد کرده‌اند. این موارد شامل مطالعات بسیاری در شبکه‌های عصبی برای حل مسائل بهینه‌سازی در زمان واقعی است ([۱۵، ۲۳، ۳۲، ۳۷، ۴۹، ۵۲، ۵۶، ۵۷، ۵۸، ۶۸، ۶۹، ۱۰۰، ۱۰۴، ۱۰۸، ۱۱۳] را ببینید).

شبکه‌ها از نیم قرن پیش از کار پیشگامان مک کلوج^۸ و پیتس^۹ متولد شدند و شبکه‌های عصبی برای اولین بار در دهه ۱۹۸۰ در زمینه بهینه‌سازی معرفی شدند. ([۳۷] و [۸۶] را ببینید). ایده اصلی روش شبکه عصبی برای بهینه‌سازی ایجاد یک تابع انرژی نامنفی و یک سیستم دینامیکی است که یک نمایه از یک شبکه عصبی مصنوعی است. سیستم دینامیکی به طور معمول در قالب معادلات دیفرانسیل معمولی است. انتظار می‌رود برای یک نقطه اولیه سیستم دینامیکی به نقطه تعادلش نزدیک شود، که جواب مساله بهینه‌سازی اصلی است. یک ضرورت اصلی این است که همزمان که یک سیستم دینامیکی به یک نقطه

^۲ Alternating direction augmented Lagrangian

^۳ Spectral bundle

^۴ Robot grasping

^۵ Large scale integration

^۶ Optical technologies.

^۷ Numerical ordinary differential equation

^۸ McCulloch

^۹ Pitts

تعادلش نزدیک می‌شود، تابع انرژی به طور یکنواخت کاهش می‌یابد. مزیت اصلی روش شبکه عصبی در بهینه‌سازی این است که ماهیت روش جواب دینامیکی به طور ذاتی موازی است. بنابراین رویکرد شبکه عصبی می‌تواند مسائل بهینه‌سازی را در زمان بسیار سریعتر از بیشتر الگوریتم‌های بهینه‌سازی محبوب که بر روی رایانه‌های دیجیتالی چند منظوره اجرا می‌شوند، حل کند. به علاوه برخی از شبکه‌های عصبی پیشنهاد شده عملکرد بهتری نسبت به دیگر شبکه‌ها از حیث نظریه یا پیاده‌سازی دارند. لازم به ذکر است که با توجه به وضعیت سیستم های نامشخص (غیر قطعی)، تجزیه و تحلیل پایداری مدل‌های شبکه عصبی می‌تواند گسترش یابد. برای اطلاعات بیشتر [۷، ۴۸، ۸۶] را ببینید.

اگرچه انواع مختلفی از شبکه‌های عصبی آنالوگ ارائه شده برای محاسبات وجود دارد، اما تنها یک گزارش از حل مسائل زمان واقعی نیمه معین با پیچیدگی محاسباتی بالا با استفاده از شبکه عصبی وجود دارد [۴۲]. ساختار این مدل شبکه عصبی پیچیده است و مدل ساده‌تری می‌تواند ارائه شود. بنابراین، ساختن یک مدل شبکه عصبی متفاوت برای مساله نیمه معین خطی با یک ساختار ساده، پیچیدگی کم، پایداری و همگرایی مناسب لازم به نظر می‌رسد. این انگیزه اصلی این فصل است. یکی دیگر از انگیزه‌های این کار این است که با پیشنهاد یک طرح بهینه‌سازی دینامیکی به حل مساله نیمه معین کمک کند. بر اساس شرایط بهینگی کارش-کان-تاکر، یک مدل شبکه عصبی کارا با تحلیل پایداری دقیق ارائه می‌شود که نقطه تعادل آن معادل نقطه‌ی کارش کان تاکر مساله نیمه معین است. با برخی مفروضات پایداری نقطه تعادل نیز به دست می‌آید.

در این فصل ما یک مدل شبکه عصبی مبتنی بر شرایط KKT ارائه خواهیم داد. ما نشان می‌دهیم که نقطه تعادل شبکه عصبی ارائه شده با نقطه کارش-کان-تاکر برنامه‌ریزی نیمه معین (۴.۲) و (۵.۲) معادل است. با به کار بردن یک تابع لیاپانوف مناسب نشان خواهیم داد که مدل شبکه عصبی ارائه شده پایدار لیاپانوف و همگرای سراسری به جواب بهینه برنامه‌ریزی نیمه معین (۴.۲) و (۵.۲) است. مثال‌های مختلف نتایج نظری را تایید می‌کنند.

۲.۳ مدل سیستم دینامیکی

با استفاده از از یک روش استاندارد بهینه‌سازی، مساله برنامه‌ریزی نیمه معین (۴.۲) و (۵.۲) را به یک سیستم دینامیکی غیرخطی تبدیل می‌کنیم. اکنون متغیرهای وابسته به زمان $x(\cdot)$ و $Z(\cdot)$ را در نظر می‌گیریم. ما یک مدل شبکه عصبی بازگشتی برای حل برنامه‌ریزی نیمه معین (۴.۲) و (۵.۲) ارائه می‌دهیم که معادله دینامیکی آن به صورت زیر داده شده است

$$\frac{dx}{dt} = -(\nabla f(x) + DA(x)^*[Z + A(x)]_+), \quad (1.3)$$

$$\frac{dZ}{dt} = [Z + A(x)]_+ - Z, \quad (2.3)$$

که $DA(x) = [A_1, A_2, \dots, A_n]$. فرض کنید $z = \text{vec}(Z)$ ، $a^i = \text{vec}(A_i)$ ، $(i = 1, \dots, n)$ ، $a(x) = \text{vec}(A(x))$ و $(z + a(x))_+ = \text{vec}([Z + A(x)]_+)$ با استفاده از (۲.۲) و (۳.۲)، شبکه عصبی (۱.۳) و (۲.۳) به صورت زیر نوشته می شود

$$\begin{aligned} \frac{dx}{dt} &= -(\nabla f(x) + \bar{A}(z + a(x))_+), \\ \frac{dz}{dt} &= (z + a(x))_+ - z. \end{aligned}$$

برای ساده شدن فرض کنید $y = (x^T, z^T)^T \in \mathbb{R}^{n+m}$ و

$$\Phi(y) = \begin{bmatrix} -(\nabla f(x) + \bar{A}(z + a(x))_+) \\ (z + a(x))_+ - z \end{bmatrix}. \quad (۳.۳)$$

بنابراین شبکه عصبی (۱.۳) و (۲.۳) به صورت زیر نوشته می شود

$$\frac{dy}{dt} = \rho \Phi(y), \quad (۴.۳)$$

$$y(t_0) = y_0, \quad \rho > 0, \quad (۵.۳)$$

که $y_0 \in S_0 = \{y = (x^T, z^T)^T \in \mathbb{R}^{n+m} \mid Z \in S_+^m \text{ \& } z = \text{vec}(Z)\}$ و ρ یک پارامتر اسکالر و نشان دهنده نرخ همگرایی شبکه عصبی (۴.۳) و (۵.۳) است. واضح است که هرچه ρ بزرگتر باشد نرخ همگرایی بهتر خواهد بود.

۳.۳ تحلیل پایداری و همگرایی

در این بخش پایداری و همگرایی شبکه عصبی پیشنهاد شده (۴.۳) و (۵.۳) ارائه می شود. برای راحتی و بدون کاستن از کلیت فرض می کنیم $\rho = 1$.

قضیه ۱.۳.۳. فرض کنید $y^* = (x^{*T}, z^{*T})^T$ یک نقطه تعادل شبکه عصبی (۴.۳) و (۵.۳) باشد، آنگاه x^* یک نقطه KKT از مسئله نیمه معین (۴.۲) و (۵.۲) است. از طرف دیگر، اگر $x^* \in \mathbb{R}^n$ یک جواب بهینه از مسئله نیمه معین (۴.۲) و (۵.۲) باشد، آنگاه یک $Z^* \in S_+^m$ وجود دارد به طوری که $y^* = (x^{*T}, z^{*T})^T$ یک نقطه تعادل شبکه عصبی (۴.۳) و (۵.۳) است که $z^* = \text{vec}(Z^*)$.

برهان. فرض کنید $y^* = (x^{*T}, z^{*T})^T$ یک نقطه تعادل شبکه عصبی (۴.۳) و (۵.۳) باشد. آنگاه $\frac{dx^*}{dt} = 0$ و $\frac{dz^*}{dt} = 0$. به راحتی نتیجه می شود

$$\nabla f(x^*) + \bar{A}(z^* + a(x^*))_+ = 0, \quad (۶.۳)$$

$$(z^* + a(x^*))_+ = z^*. \quad (۷.۳)$$

واضح است که $(z^* + a(x^*))_+ = z^*$ معادل است با $[Z^* + A(x^*)]_+ = Z^*$. بنابراین از لم ۲.۲.۲ داریم

$$Z^* \succeq \circ, \quad A(x^*) \preceq \circ, \quad \langle Z^*, A(x^*) \rangle = \circ. \quad (۸.۳)$$

با استفاده از (۳.۲) داریم $DA(x^*)^* Z^* = \bar{A} z^*$ و با جایگذاری (۷.۳) در (۶.۳) داریم

$$\nabla f(x^*) + DA(x^*)^* Z^* = \circ. \quad (۹.۳)$$

از (۸.۳) و (۹.۳)، می‌بینیم که $y^* = (x^{*T}, z^{*T})^T$ در شرایط بهینگی (۲۵.۲)–(۲۷.۲) صدق می‌کند. اثبات عکس سراسر است. \square

قضیه ۲.۳.۳. برای هر نقطه شروع $y_0 = y(t_0) \in S_0$ ، یک جواب پیوسته برای سیستم (۴.۳) و (۵.۳) وجود دارد.

برهان. چون $\nabla f(x)$ و $DA(x)$ روی مجموعه محدب باز $B \subseteq \mathbb{R}^{n+m}$ پیوسته هستند بنابراین $\nabla f(x) + \bar{A}(z + a(x))_+ - z$ و $(z + a(x))_+ - z$ پیوسته لیپ شیتز محلی هستند. طبق قضیه وجود جواب محلی برای معادلات دیفرانسیل معمولی ۵.۳.۱، شبکه عصبی (۴.۳) و (۵.۳) یک جواب پیوسته منحصر به فرد $y(t)$ برای $t \in [t_0, \tau)$ ، $\tau > t_0$ دارد. \square

قضیه ۳.۳.۳. جواب شبکه عصبی (۴.۳) و (۵.۳) با نقطه شروع $y_0 \in S_0$ پایدار لیپانوف است. علاوه بر این، مسیر جواب شبکه عصبی (۴.۳) و (۵.۳) جواب ماکزیمال نیست.

برهان. قرار دهید $\bar{z} = (z + a(x))_+$ و $\bar{x} = x - \nabla f(x) - \bar{A}\bar{z}$ و تابع لیپانوف زیر را در نظر بگیرید

$$\begin{aligned} E(x, z) &= f(x) - f(x^*) + \frac{1}{\rho} [\|\bar{z}\|^2 - \|z^*\|^2] - (x - x^*)^T (\nabla f(x^*) + \bar{A}z^*) \\ &\quad - (z - z^*)^T z^* + \frac{1}{\rho} \|x - x^*\|^2 + \frac{1}{\rho} \|z - z^*\|^2, \end{aligned} \quad (۱۰.۳)$$

که $y^* = (x^{*T}, z^{*T})^T$ یک نقطه تعادل شبکه عصبی (۴.۳) و (۵.۳) است. آنگاه

$$\begin{aligned}
 \frac{dE(y(t))}{dt} &= [\nabla f(x) + \bar{A}\bar{z} - \nabla f(x^*) - \bar{A}z^* + x - x^*]^T (\bar{x} - x) \\
 &+ (\bar{z} + z - \bar{z}z^*)^T (\bar{z} - z) / \gamma \\
 &= -\|x - \bar{x}\|^2 + [\nabla f(x) + \bar{A}\bar{z} - \nabla f(x^*) - \bar{A}z^* + \bar{x} - x^*]^T (\bar{x} - x) \\
 &- \frac{1}{\gamma} \|z - \bar{z}\|^2 + (\bar{z} - z^*)^T (\bar{z} - z) \\
 &= -\|x - \bar{x}\|^2 - [\nabla f(x) - \nabla f(x^*)]^T (x - x^*) - (\nabla f(x^*) + \bar{A}z^*)^T (\bar{x} - x^*) \\
 &[x - \nabla f(x) - \bar{A}\bar{z} - \bar{x}]^T (\bar{x} - x^*) - (\bar{A}\bar{z} - \bar{A}z^*)^T (x - x^*) \\
 &- \frac{1}{\gamma} \|z - \bar{z}\|^2 + (\bar{z} - z^*)^T (a(x) - (z + a(x))_-) \\
 &= -\|x - \bar{x}\|^2 - \frac{1}{\gamma} \|z - \bar{z}\|^2 \\
 &- (\nabla f(x^*) + \bar{A}z^*)^T (\bar{x} - x^*) - [x - \nabla f(x) - \bar{A}\bar{z} - \bar{x}]^T (\bar{x} - x^*) \\
 &- \bar{z}^T [-\bar{A}^T(x^* - x) - a(x) + a(x^*)] + \bar{z}^T a(x^*) - \bar{z}(z + a(x))_- \\
 &+ z^{*T} [\bar{A}^T(x - x^*) - a(x) + a(x^*)] - z^* a(x^*) + z^*(z + a(x))_- \\
 &= -\|x - \bar{x}\|^2 - \frac{1}{\gamma} \|z - \bar{z}\|^2 \\
 &- (\nabla f(x^*) + \bar{A}z^*)^T (\bar{x} - x^*) - [x - \nabla f(x) - \bar{A}\bar{z} - \bar{x}]^T (\bar{x} - x^*) \\
 &- \bar{z}^T [-\bar{A}^T(x^* - x) - a(x) + a(x^*)] + \bar{z}^T a(x^*) \\
 &- z^{*T} [-\bar{A}^T(x - x^*) + a(x) - a(x^*)] + z^*(z + a(x))_- \\
 &\leq -\|x - \bar{x}\|^2 - \frac{1}{\gamma} \|z - \bar{z}\|^2 \\
 &- (\nabla f(x^*) + \bar{A}z^*)^T (\bar{x} - x^*) - [x - \nabla f(x) - \bar{A}\bar{z} - \bar{x}]^T (\bar{x} - x^*) \\
 &- \bar{z}^T [-\bar{A}^T(x^* - x) - a(x) + a(x^*)] \\
 &- z^{*T} [-\bar{A}^T(x - x^*) + a(x) - a(x^*)]. \tag{۱۱.۳}
 \end{aligned}$$

قرار دهید $u = x^*$ و $v = x - \nabla f(x) - \bar{A}\bar{z}$ ، با استفاده از لم ۲ در [۱۰۹] داریم

$$(x - \nabla f(x) - \bar{A}\bar{z} - \bar{x})^T (\bar{x} - x^*) \geq 0. \tag{۱۲.۳}$$

از محدب و مشتق پذیر بودن $A(x)$ داریم

$$\begin{cases} A(x^*) \preceq A(x) + DA(x)^*(x^* - x), \\ A(x) \preceq A(x^*) + DA(x^*)^*(x - x^*). \end{cases}$$

از آنجایی که $\bar{z}, z^* \in S_+^m$ ، با استفاده از لم ۲.۲.۲ داریم

$$\begin{cases} \bar{z}^T [a(x^*) - a(x) + \bar{A}^T(x^* - x)] \leq 0, \\ z^{*T} [a(x) - a(x^*) - \bar{A}^T(x - x^*)] \leq 0. \end{cases} \tag{۱۳.۳}$$

به‌علاوه چون y^* یک نقطه تعادل شبکه عصبی (۴.۳) و (۵.۳) است، $\phi(y^*) = 0$ و در نتیجه

$$\nabla f(x^*) + \bar{A}z^* = 0. \quad (14.3)$$

با جایگذاری (۱۲.۳) - (۱۴.۳) در (۱۱.۳) نتیجه می‌گیریم

$$\frac{dE(y(t))}{dt} \leq -\|x - \bar{x}\|^2 - \frac{1}{\rho}\|z - \bar{z}\|^2 \leq 0. \quad (15.3)$$

این یعنی اینکه شبکه عصبی (۴.۳) و (۵.۳) پایدار لیپانوف است. علاوه بر این

$$E(y_0) \geq E(y(t)) \geq \frac{1}{\rho}\|y(t) - y^*\|^2, \quad (16.3)$$

نشان می‌دهد که مسیر جواب $y(t)$ کراندار است و بنابراین قضیه ۶.۳.۱ نتیجه می‌شود $y(t)$ ،
 \square $t \in [t_0, \tau)$ وقتی $\tau \rightarrow +\infty$ وجود دارد.

قضیه ۴.۳.۳. فرض کنید y^* نقطه تعادل شبکه عصبی (۴.۳) و (۵.۳) باشد. جواب شبکه عصبی (۴.۳) و (۵.۳) با نقطه آغازین $y_0 \in S_0$ همگرای سراسری به جواب بهینه مسئله نیمه معین (۴.۲) و (۵.۲) است.

برهان. از (۱۶.۳) $y(t)$ کراندار است، بنابراین یک نقطه حدی $\hat{y} = (\hat{x}^T, \hat{z}^T)^T$ و یک دنباله نزولی $\{t_k\}$ ($t_k \rightarrow +\infty$) وجود دارد به طوری که

$$\lim_{k \rightarrow \infty} y(t_k) = \hat{y}. \quad (17.3)$$

از (۱۶.۳)، وقتی که $\|y(t)\| \rightarrow \infty$ داریم $E(y) \rightarrow \infty$. پس مجموعه سطح

$$M(y_0) = \{y \in S_0 \mid E(y) \leq E(y_0)\}$$

کراندار است. همچنین $M(y_0)$ زیر مجموعه بسته \mathbb{R}^{n+m} و در نتیجه فشرده است. توجه داشته باشید که $\frac{dE(y(t))}{dt} \leq 0$ باعث می‌شود $y(t) \in M(y_0)$ برای هر $t \geq t_0$. با بکار بردن اصل تغییرناپذیری لسال در (۱۰.۳.۱) درمی‌یابیم که وقتی $t \rightarrow \infty$ ، $\{y(t) \rightarrow M\}$ که در آن M بزرگترین مجموعه تغییرناپذیر در $\{y \in M(y_0) \mid \frac{dE(y(t))}{dt} = 0\}$ است. بنابراین از (۱۷.۳) داریم $\hat{y} \in M$. از (۱۵.۳) نتیجه می‌شود $\frac{dE(y(t))}{dt} = 0$ اگر و فقط اگر $\frac{dx}{dt} = 0$ و $\frac{dz}{dt} = 0$. پس \hat{y} یک نقطه تعادل شبکه عصبی (۴.۳) و (۵.۳) است. با استفاده از قضیه ۱.۳.۳، $\hat{y} = (\hat{x}^T, \hat{z}^T)^T$ ، نقطه KKT مسئله نیمه معین (۴.۲) و (۵.۲) است.

با جایگذاری $x^* = \hat{x}$ و $z^* = \hat{z}$ در (۱۰.۳) تابع لیپانوف دیگری به شکل زیر تعریف می‌کنیم

$$\begin{aligned} \hat{E}(x, z) &= f(x) - f(\hat{x}) + \frac{1}{\rho}[\|z\|^2 - \|\hat{z}\|^2] - (x - \hat{x})^T (\nabla f(\hat{x}) + \bar{A}\hat{z}) \\ &\quad - (z - \hat{z})^T \hat{z} + \frac{1}{\rho}\|x - \hat{x}\|^2 + \frac{1}{\rho}\|z - \hat{z}\|^2. \end{aligned}$$

$\hat{E}(x, z)$ پیوسته مشتق پذیر بوده و $\hat{E}(\hat{x}, \hat{z}) = 0$. بنابراین، از (۱۷.۳)

$$\hat{E}_{k \rightarrow \infty}(x_k, z_k) = \hat{E}(\hat{x}, \hat{z}) = 0.$$

پس برای هر $\epsilon > 0$ یک $q > 0$ وجود دارد به طوری که برای تمام $t > t_q$ ، $\frac{d\hat{E}(x, z)}{dt} \leq 0$. مشابه فرایند بالا می توانیم اثبات کنیم که $\frac{d\hat{E}(x, z)}{dt} \leq 0$. از اینکه $t \geq t_q$ نتیجه می شود

$$\frac{1}{q} (\|x(t) - \hat{x}\|^2 + \|z(t) - \hat{z}\|^2) \leq \hat{E}(x, z) < \epsilon.$$

یعنی $\lim_{t \rightarrow \infty} x(t) = \hat{x}$ و $\lim_{t \rightarrow \infty} z(t) = \hat{z}$. بنابراین منحنی جواب شبکه عصبی (۴.۳) و (۵.۳) همگرایی سراسری به یک نقطه بهینگی مسئله نیمه معین (۴.۲) و (۵.۲) است. □

نتیجه ۱.۳.۳. فرض کنید $D^* = \{(x^*, Z^*)\}$ مجموعه تمام نقاط بهینگی مسئله نیمه معین (۴.۲) و (۵.۲) باشد. آنگاه شبکه عصبی (۴.۳) و (۵.۳) پایدار جانبی سراسری به نقطه تعادل $y^* = (x^{*T}, z^{*T})^T$ است که $z^* = \text{vec}(Z^*)$

۴.۳ مثال های عددی

در این بخش برای نشان دادن اعتبار مدل شبکه عصبی ارائه شده چندین مثال عددی آورده می شود. برای این مثال ها همچنین کارایی عددی شبکه عصبی (۴.۳) و (۵.۳) را با مقادیر مختلف ρ و نقاط آغازین مختلف y_0 بررسی می کنیم. شبیه سازی این مثال ها با استفاده از نرم افزار Matlab 7 و حل کننده معادلات دیفرانسیل معمولی ode45 انجام شده است.

مثال ۱.۳ [۴۲] مساله نیمه معین (۴.۲) و (۵.۲) را در نظر بگیرید که

$$m = 4, \quad n = 4, \quad c = \begin{bmatrix} 4/7072 & -6/4875 & -31/5557 & 31/5310 \end{bmatrix}^T,$$

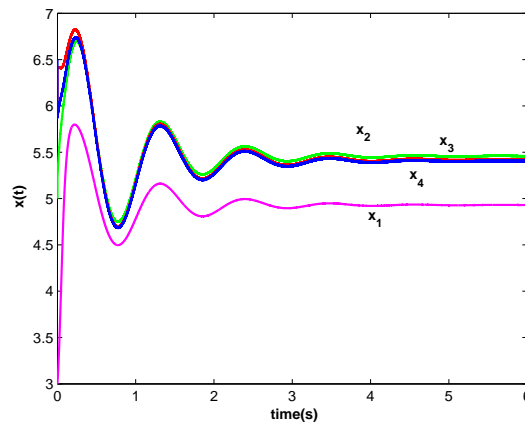
$$A_0 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -5 \end{pmatrix}, \quad A_1 = \begin{pmatrix} -0/2786 & -0/4406 & 0/8163 & 0 \\ -0/4406 & -0/7472 & -0/2046 & 0 \\ 0/8163 & -0/2046 & 0/258 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$A_2 = \begin{pmatrix} -1/6951 & 0/2024 & 1/0054 & 0 \\ 0/2024 & 5/0589 & 0/9851 & 0 \\ 1/0054 & 0/9851 & -4/3638 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$A_3 = \begin{pmatrix} -3/2736 & -6/0849 & -0/2810 & 0 \\ -6/0849 & 8/6117 & -0/3530 & 0 \\ -1/9324 & -0/3772 & 10/2313 & 0 \\ 0 & 0 & 0 & -2 \end{pmatrix},$$

$$A_4 = \begin{pmatrix} 4/8445 & 6/3682 & -1/9324 & 0 \\ 6/3682 & -13/0758 & -0/3772 & 0 \\ -1/9324 & -0/3772 & 10/2313 & 0 \\ 0 & 0 & 0 & -2 \end{pmatrix}.$$

این مساله دارای جواب بهینه $(4/9299, 5/4578, 5/4177, 5/4038)^T$ است. شکل ۱.۳ رفتار منحنی های $(x_1(t), x_2(t), x_3(t), x_4(t))^T$ مبتنی بر شبکه عصبی (۴.۳) و (۵.۳) را با نقاط آغازین تصادفی $x_0 \in S_+^m$ و Z_0 نشان می دهد.



شکل ۱.۳: رفتار همگرایی $(x_1(t), x_2(t), x_3(t), x_4(t))^T$ شبکه عصبی (۴.۳) و (۵.۳) در مثال ۱.۳.

مثال ۲.۳ [۸] مساله برنامه ریزی خطی زیر را در نظر بگیرید

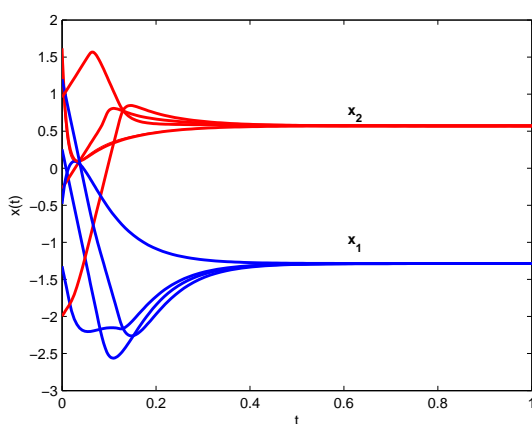
$$\begin{aligned} & \text{minimize} && 3x_1 - x_2 \\ & \text{subject to} && x_1 - 3x_2 \geq -3, \\ & && 2x_1 + x_2 \geq -2, \\ & && 2x_1 + x_2 \leq 8, \\ & && 4x_1 - x_2 \leq 16. \end{aligned}$$

جواب بهینه این مساله $x_1 = \frac{5}{4}$ و $x_2 = -\frac{19}{4}$ است. همانطور که در فصل دوم بحث شد این مساله می تواند به عنوان مسئله نیمه معین (۴.۲) و (۵.۲) طرح شود که

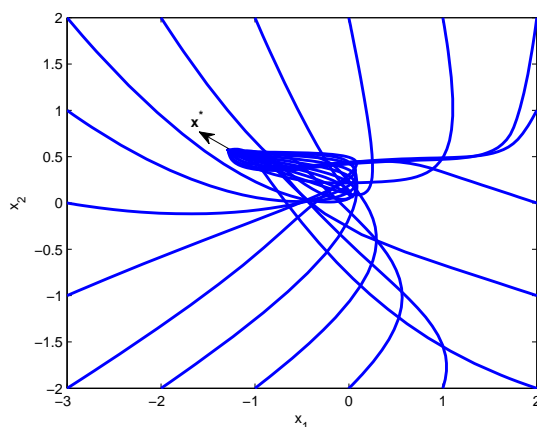
$$A_0 = \text{diag}(3, 2, 8, 16), \quad A_1 = \text{diag}(-1, -2, 2, 4), \quad A_2 = \text{diag}(3, -1, 1, -1).$$

شکل های ۲.۳ و ۳.۳ به ترتیب مسیره های گذرا و منحنی های صفحه فاز متغیرهای وضعیت $x_1(t)$ و $x_2(t)$ را نشان می دهند.

مثال ۳.۳ [۲۵] مساله مینیمم کردن بزرگترین مقدار ویژه ماتریس $A(x) = A_0 + \sum_{i=1}^5 x_i A_i$ را در نظر بگیرید که در آن



شکل ۲.۳: رفتار همگرایی $(x_1(t), x_2(t))^T$ از شبکه عصبی (۴.۳) و (۵.۳) با ۴ نقطه آغازین تصادفی در مثال ۲.۳



شکل ۳.۳: نمودار فاز $(x_1(t), x_2(t))^T$ از شبکه عصبی (۴.۳) و (۵.۳) با ۱۸ نقطه آغازین در مثال ۲.۳.

$$A_0 = \begin{pmatrix} -0.69 & -0.32 & 0.34 & 0.43 & -0.05 \\ -0.32 & -0.11 & -0.11 & -0.45 & -0.34 \\ 0.34 & -0.11 & -0.71 & -0.33 & -0.08 \\ 0.43 & -0.45 & -0.33 & 0.65 & 0.27 \\ -0.05 & -0.34 & -0.08 & 0.27 & 0.39 \end{pmatrix}$$

$$A_1 = \begin{pmatrix} -0/66 & 0/31 & 0/57 & -0/06 & -0/44 \\ 0/31 & -0/23 & -0/12 & -0/35 & 0/28 \\ 0/57 & -0/12 & -0/26 & -0/06 & -0/37 \\ -0/06 & -0/35 & -0/06 & 0/64 & 0/34 \\ -0/44 & 0/28 & -0/37 & 0/34 & 0/61 \end{pmatrix},$$

$$A_2 = \begin{pmatrix} -0/31 & 0/35 & 0/06 & -0/23 & 0/17 \\ 0/35 & 0/24 & -0/19 & 0/21 & -0/12 \\ 0/06 & -0/19 & -0/34 & 0/00 & 0/36 \\ -0/23 & 0/21 & 0/00 & 0/16 & -0/24 \\ 0/17 & -0/12 & -0/36 & -0/24 & 0/00 \end{pmatrix},$$

$$A_3 = \begin{pmatrix} 0/27 & -0/14 & 0/13 & -0/32 & -0/08 \\ -0/14 & -0/20 & -0/29 & -0/05 & -0/64 \\ 0/13 & -0/29 & -0/45 & -0/20 & -0/59 \\ -0/32 & -0/05 & -0/20 & -0/27 & -0/46 \\ -0/08 & -0/64 & -0/59 & -0/46 & -0/39 \end{pmatrix},$$

$$A_4 = \begin{pmatrix} -0/57 & -0/38 & -0/09 & 0/31 & 0/22 \\ -0/38 & 0/66 & 0/17 & -0/03 & 0/51 \\ -0/09 & 0/17 & 0/23 & 0/12 & -0/21 \\ 0/31 & -0/03 & 0/12 & -0/56 & -0/21 \\ 0/22 & 0/51 & -0/21 & -0/21 & 0/59 \end{pmatrix},$$

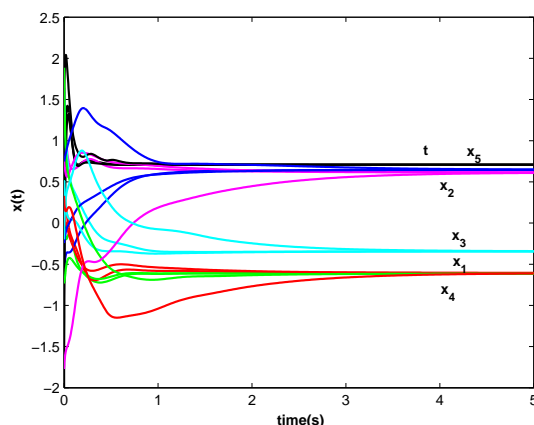
$$A_5 = \begin{pmatrix} 0/22 & 0/28 & 0/14 & 0/03 & 0/09 \\ 0/28 & 0/69 & -0/12 & 0/10 & 0/30 \\ 0/14 & -0/12 & -0/77 & -0/21 & 0/13 \\ 0/03 & 0/10 & -0/21 & -0/42 & -0/15 \\ 0/09 & 0/30 & 0/13 & -0/15 & 0/22 \end{pmatrix}.$$

در فصل دوم توضیح دادیم که این مساله می تواند به عنوان یک مسئله نیمه معین به صورت زیر ارائه شود

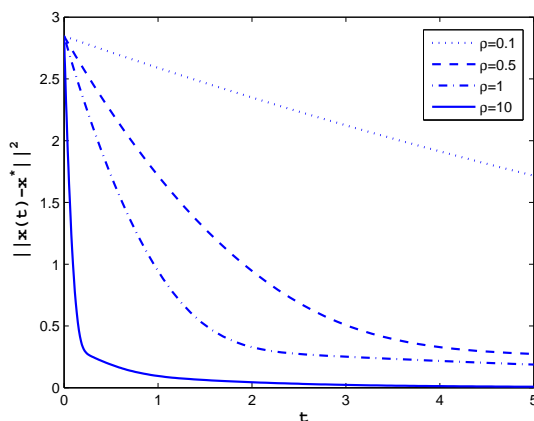
$$\begin{aligned} & \text{minimize} \quad \lambda \\ & \text{subject to} \quad B(\lambda, x) = \lambda I - A(x) \succeq 0. \end{aligned}$$

جواب بهینه این مساله $x^* = (-0/6136, 0/6145, -0/3437, -0/6067, 0/6465)^T$ و مقدار بهینه آن $\lambda^* = 0/7089$ است. شکل ۴.۳ نشان می دهد که مسیرهای شبکه عصبی (۴.۳) و (۵.۳) برای حل این مساله با ۵ نقطه شروع تصادفی به جواب بهینه مساله همگراست.

در این مساله تاثیر پارامتر ρ روی $\|x(t) - x^*\|^2$ را در شبکه عصبی (۴.۳) و (۵.۳) بررسی می‌کنیم. شکل ۵.۳ رفتار نرم خطای $\|x(t) - x^*\|^2$ را با ۴ مقدار مختلف برای ρ و نقطه شروع $(\lambda_0, x_0^T)^T = (2, -1, 1, -1, 0, 1)^T$ نشان می‌دهد. مشاهده می‌کنیم که وقتی $\rho = 0.1$ شبکه عصبی (۴.۳) و (۵.۳) سرعت همگرایی کمتری نسبت به زمانی که $\rho = 10$ دارد.



شکل ۴.۳: رفتار همگرایی $(\lambda(t), x(t)^T)^T$ از شبکه عصبی (۴.۳) و (۵.۳) با ۵ نقطه آغازین تصادفی در مثال ۳.۳.

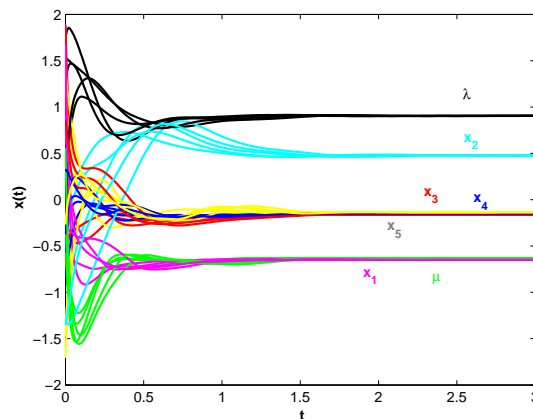


شکل ۵.۳: رفتار همگرایی $\|(\lambda(t), x(t)) - (\lambda^*, x^*)\|^2$ با $(\lambda_0, x_0^T)^T = (2, -1, 1, -1, 0, 1)^T$ در مثال ۳.۳.

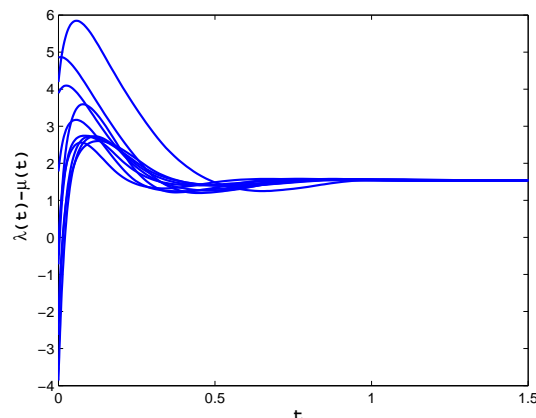
مثال ۴.۳. ماتریس‌های A_i ، $i = 0, 1, \dots, 5$ ، در مثال ۳.۳ را در نظر بگیرید. مسئله مینیمم کردن فاصله بین بزرگترین و کوچکترین مقدار ویژه ماتریس $A(x)$ است. این مساله می‌تواند به صورت مساله نیمه معین زیر طرح شود [۳۰]

$$\begin{aligned} & \text{minimize } \lambda - \mu \\ & \text{subject to } \mu I \preceq A(x) \preceq \lambda I, \end{aligned}$$

که $\lambda = \lambda_{max}(A(x))$ و $\mu = \lambda_{min}(A(x))$ به ترتیب بزرگترین و کوچکترین مقدار ویژه ماتریس $A(x)$ هستند. تمام نتایج عددی نشان می‌دهند که شبکه عصبی (۴.۳) و (۵.۳) به جواب بهینه $(\lambda^*, \mu^*, x^{*T})^T$ همگراست و مقدار بهینه تابع هدف $1/5389$ است. شکل‌های ۶.۳ و ۷.۳ رفتار مسیرهای گذرا $(\lambda(t), \mu(t), x(t)^T)^T$ و تابع هدف را به ترتیب با ۵ و ۱۰ نقطه آغازین تصادفی نشان می‌دهند. دوباره، همگرایی سریع به جواب بهینه به خوبی مشاهده می‌شود. شکل ۸.۳ تاثیر پارامتر ρ را بر مقدار $\|x(t) - x^*\|^2$ نشان می‌دهد. از شکل ۸.۳ ما می‌بینیم زمانی که $\rho = 0.5$ شبکه عصبی (۴.۳) و (۵.۳) سرعت همگرایی کمتری از $\|x(t) - x^*\|^2$ را نسبت به زمانی که $\rho = 20$ دارد.

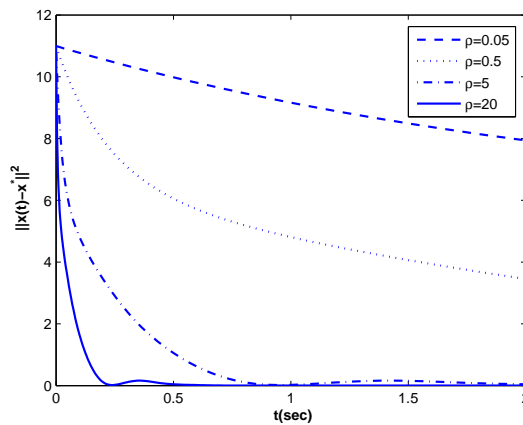


شکل ۶.۳: رفتار همگرایی $(\lambda, \mu, x^T)^T$ شبکه عصبی (۴.۳) و (۵.۳) با ۵ نقطه آغازین تصادفی در مثال ۴.۳.



شکل ۷.۳: رفتار همگرایی تابع هدف با ۱۰ نقطه آغازین تصادفی در مثال ۴.۳.

مثال ۵.۳. مسئله مینیمم کردن $\|A(x)\|$ ، نرم ماتریس $A(x) = A_0 + x_1 A_1 + \dots + x_5 A_5$ است که $A_i, i = 0, 1, \dots, 5$ در مثال ۳.۳ داده شده‌اند. این مسئله می‌تواند به صورت مسئله نیمه معین زیر طرح شود [۸۸]



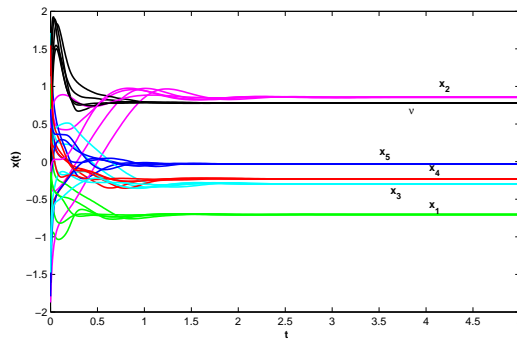
شکل ۸.۳: رفتار همگرایی $\|x(t) - x^*\|^2$ با $(\lambda_0, \mu_0, x_0^T)^T = (1, -1, 1, -1, 0, 2, 1)^T$ در مثال ۴.۳.

minimize ν

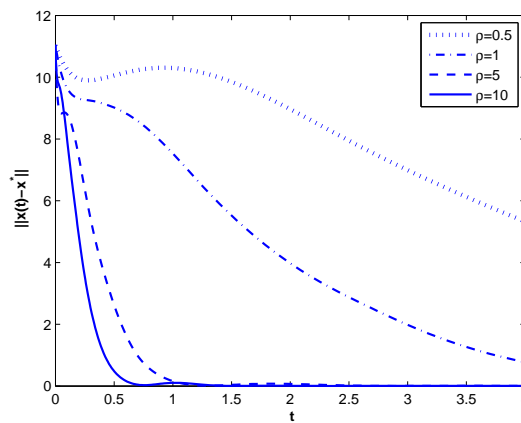
$$\text{subject to } B(\nu, x) = \begin{pmatrix} \nu I & A(x) \\ A(x)^T & \nu I \end{pmatrix} \succeq 0.$$

ما شبکه عصبی (۴.۳) و (۵.۳) را برای حل این مسئله به کار می‌گیریم. تمام نتایج نشان می‌دهند که شبکه عصبی (۴.۳) و (۵.۳) به جواب بهینه $(\nu^*, x^{*T})^T$ همگراست و مقدار بهینه تابع هدف 0.7806 است. برای مثال، شکل ۹.۳ منحنی‌های متغیرهای وضعیت $(\nu, x(t)^T)^T$ را با ۵ نقطه آغازین تصادفی نشان می‌دهد. همگرایی سریع شبکه عصبی مشاهده می‌شود. شکل ۱۰.۳ رفتار همگرایی منحنی خطا را با مقادیر مختلف ρ و با نقطه آغازین $(\nu_0, x_0^T)^T = (2, 1, 2, -2, 1, 1)^T$ نشان می‌دهد. واضح است که ρ بزرگتر، نرخ همگرایی بهتر را نتیجه می‌دهد. در این مثال ما شرط مکمل (۲۶.۲) برای مسئله نیمه معین بررسی می‌کنیم که $\|Z^* B(\nu^*, x^*)\| = 0.000008$ است.

مثال ۶.۳. [۲۵] مسئله نیمه معین (۴.۲) و (۵.۲) را در نظر بگیرید که $A_1 = (A_{ij}^1)$ ، سایر مولفه های تمام این ماتریس ها برابر صفر و $c = (1, 1, 1)^T$ است. جواب بهینه این مسئله $(2, -1, 0)^T$ است. شکل ۱۱.۳ نشان می‌دهد که شبکه عصبی (۴.۳) و (۵.۳) برای حل این مساله با ۵ نقطه شروع تصادفی به جواب بهینه مساله همگراست. شکل ۱۲.۳ رفتار همگرایی متغیر دوگان $Z(t)$ در شبکه عصبی (۴.۳) و (۵.۳) را با نقاط آغازین اولیه و دوگان زیر نشان می‌دهد



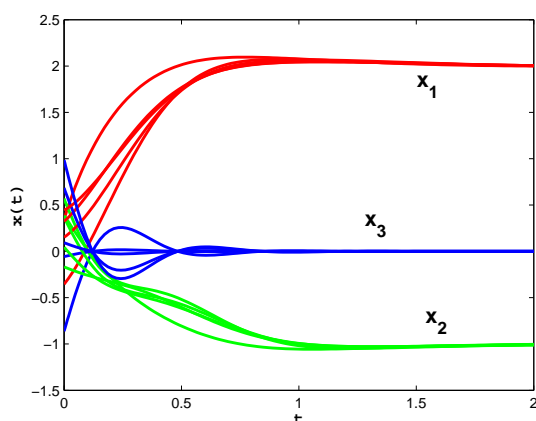
شکل ۹.۳: رفتار همگرایی $(\nu(t), x(t)^T)^T$ در شبکه عصبی (۴.۳) و (۵.۳) با ۵ نقطه آغازین تصادفی در مثال ۵.۳



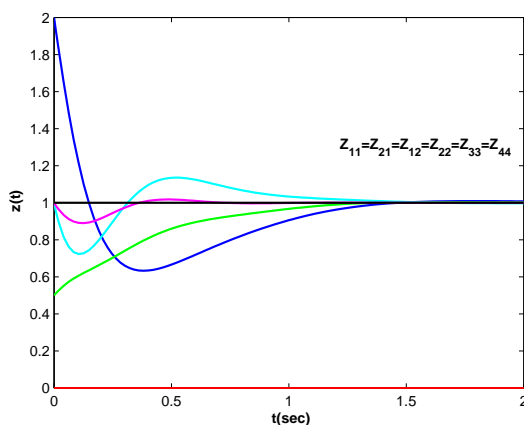
شکل ۱۰.۳: رفتار همگرایی $\|(\nu(t), x(t)^T)^T - (\nu^*, x^{*T})^T\|^2$ با $(\nu_0, x_0^T)^T = (2, 1, 2, -2, 1, 1)^T$ در مثال ۵.۳

$$Z_0 = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad x_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

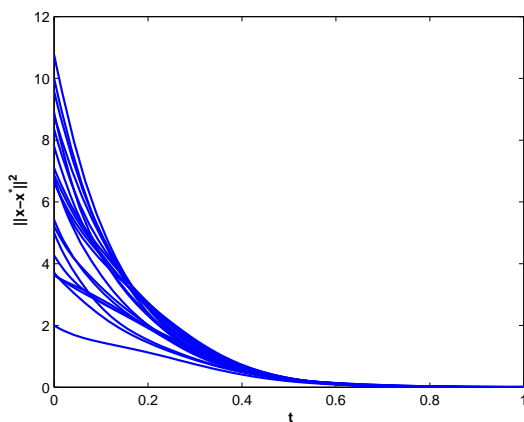
همچنین شکل ۱۳.۳ رفتار همگرایی نرم خطا، یعنی $\|x(t) - x^*\|^2$ را بر مبنای شبکه عصبی (۴.۳) و (۵.۳) با ۲۰ نقطه تصادفی و $\rho = 10$ نشان می دهد.



شکل ۱۱.۳: رفتار همگرایی $(x_1(t), x_2(t), x_3(t))^T$ از شبکه عصبی (۴.۳) و (۵.۳) با ۵ نقطه آغازین تصادفی در مثال ۶.۳.



شکل ۱۲.۳: رفتار همگرایی $Z(t)$ از شبکه عصبی (۴.۳) و (۵.۳) در مثال ۶.۳.



شکل ۱۳.۳: رفتار همگرایی $\|x(t) - x^*\|^2$ مبتنی بر شبکه عصبی (۴.۳) و (۵.۳) با ۲۰ نقطه آغازین تصادفی و $\rho = 10$ در مثال ۶.۳

فصل ۴

یک مدل شبکه عصبی گرادینانی برای حل مسئله نیمه معین خطی

۱.۴ مقدمه

همانگونه که در فصل قبل گفته شد مساله نیمه معین در سال‌های اخیر بسیار مورد توجه قرار گرفته و کاربردهای گسترده‌ای در مهندسی و بهینه‌سازی ساختاری یافته است. به تازگی تحقیقات گسترده‌ای در مورد استفاده از توابع شایسته^۱ برای حل مسئله مکمل خطی و مسئله مکمل غیرخطی انجام شده است. به این ترتیب روش‌هایی برای حل آن‌ها توسعه یافتند. به نظر می‌رسد روش‌های نیوتن بر اساس تابع فیشر برمیچر^۲ از روش‌های نقطه داخلی موثرتر باشند. یکی از رویکردهای جالب برای حل کردن بهینه‌سازی محدب غیرخطی، فرمول‌بندی آن به عنوان یک مسئله بهینه‌سازی بدون محدودیت با استفاده از یک تابع شایسته است. تابع شایسته تابعی است که مینیمم‌کننده سراسری آن جواب بهینه مساله مکمل غیرخطی است. برای نمونه در [۷۰] یک مدل شبکه عصبی مبتنی بر تابع فیشر-برمیچر برای حل مسئله مکمل درجه دوم ارائه شده است. تیسنگ در [۸۷] توسعه توابع شایسته مورد استفاده برای مسائل مکمل خطی و غیرخطی به مسئله نیمه معین متناظرشان را مطالعه کرده

^۱Merit functions

^۲Fischer-Burmeister

است به طور ویژه تابع شایسته مبتنی بر تابع فیشر-برمیچر تعمیم یافته. با الهام گرفتن از این موضوع، در این فصل برای اولین بار یک مدل شبکه عصبی بر اساس تابع فیشر-برمیچر تعمیم یافته ارائه می‌دهیم. ثابت می‌کنیم که نقطه تعادل شبکه عصبی پیشنهادی با نقطه بهینگی کارش کان تا کر مسئله نیمه معین معادل است. همچنین پایداری و همگرایی سراسری شبکه عصبی پیشنهاد شده نیز به دست می‌آید.

۲.۴ یک مدل شبکه عصبی گرادینانی

شرایط بهینگی KKT مسئله نیمه معین در (۲۵.۲)–(۲۷.۲) را در نظر بگیرید. در این بخش برای یافتن نقطه بهینگی KKT یک مدل شبکه عصبی گرادینانی مبتنی بر تابع فیشربرمیچر پیشنهاد می‌کنیم. برای ارائه مدل شبکه عصبی مورد نظر، نگاشت $\phi: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ را به صورت زیر تعریف می‌کنیم

$$\phi(a, b) := \sqrt{a^2 + b^2} - (a + b).$$

این نگاشت توسط فیشر [۲۸] معرفی شد و به تابع فیشر-برمیچر معروف است. به راحتی می‌توان نشان داد که این تابع دارای خاصیت مهم زیر است

$$\phi(a, b) = 0 \iff a \geq 0, b \geq 0, ab = 0.$$

تعریف ۱.۲.۴. فرض کنید $A \in S_+^m$ ، آنگاه $A^\dagger \in S_+^m$ را ماتریسی تعریف می‌کنیم که $A = A^\dagger A^\dagger$. توجه کنید که $A^\dagger = P^T \text{diag}[\lambda_1^\dagger, \dots, \lambda_m^\dagger] P$ ، که در آن λ_j نشان‌دهنده j -امین مقدار ویژه ماتریس A است، وقتی که مقادیر ویژه به صورت نازولی مرتب شده‌اند. همچنین P ماتریسی است که در تجزیه طیفی ماتریس A ظاهر شده است، یعنی $A = P^T \text{diag}[\lambda_1, \dots, \lambda_m] P$.

اکنون نگاشت $\Phi: S^m \times S^m \rightarrow S^m$ را به صورت زیر تعریف می‌کنیم

$$\Phi(X, Y) = (X^2 + Y^2)^\dagger - (X + Y). \quad (1.4)$$

Φ به وضوح تعمیم تابع فیشر-برمیچر ϕ است که آرگومان‌های آن به جای اعداد حقیقی ماتریس‌های متقارن هستند. این نگاشت نه تنها از نظر ظاهری بلکه از نظر خواص نیز مشابه تابع فیشر است. لم زیر را ببینید.

لم ۱.۲.۴. تیسنگ [۸۷]، لم ۱.۶

$$\Phi(X, Y) = 0 \iff X \in S_+^m, Y \in S_+^m, XY = 0.$$

با استفاده از دو لم (۱.۲.۴) و (۳.۲.۲) شرایط بهینگی (۲۵.۲)–(۲۷.۲) را به صورت $\Phi(A(x), Z) = 0$ بازنویسی می‌کنیم. اکنون می‌خواهیم تعریف Φ را اصلاح کنیم. به این

منظور، فرض کنید μ یک عدد حقیقی باشد. تابع $\phi_\mu : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ معروف به تابع فیشر برمیچر هموار را به صورت زیر تعریف می‌کنیم

$$\phi_\mu(a, b) := \sqrt{a^2 + b^2 + 2\mu^2} - (a + b).$$

دلیل اینکه این تابع فیشر برمیچر هموار نامیده می‌شود این است که برای هر $\mu \neq 0$ به وضوح تابعی به طور پیوسته مشتق‌پذیر^۳ است و اگر $\mu = 0$ این تابع همان تابع فیشر برمیچر است. حال تابع فیشر برمیچر هموار تعمیم یافته^۴ $\Phi_\mu : S^m \times S^m \rightarrow S^m$ را به صورت زیر تعریف می‌کنیم

$$\Phi_\mu(X, Y) = (X^2 + Y^2 + 2\mu^2 I)^{\frac{1}{2}} - (X + Y). \quad (2.4)$$

سپس نتیجه زیر را داریم

لم ۲.۲.۴ [۴۶] فرض کنید $\mu > 0$ یک عدد دلخواه و Φ_μ به صورت (۲.۴) تعریف شده باشد، آنگاه

$$\Phi_\mu(X, Y) = 0 \iff X \in S_+^m, Y \in S_+^m, XY = \mu^2 I. \quad (3.4)$$

تعریف ۲.۲.۴ [۳۸] برای هر $C \in S_+^m$ نگاشت خطی $\mathcal{L}_C : S^m \rightarrow S^m$ را به صورت زیر تعریف می‌کنیم

$$\mathcal{L}_C[X] := CX + XC. \quad (4.4)$$

\mathcal{L}_C به طور اکید یکنوا^۵ است (زیرا اگر $X \neq 0$ آنگاه $\langle x, \mathcal{L}_C[X] \rangle = 2tr[CX^2] > 0$) بنابراین معکوس‌پذیر است، یعنی برای هر $X \in S_m$ ، $\mathcal{L}_C^{-1}[X]$ ماتریسی منحصر به فرد مثل $D \in S^m$ است به طوری که در $CD + DC = X$ صدق می‌کند ([۳۸]، قضیه ۲.۲.۳) را ببینید). به علاوه $\mathcal{L}_C^{-1}[X]$ در (X, C) پیوسته است.

لم ۳.۲.۴ [۴۶] اگر Φ_μ به صورت (۲.۴) بوده و $X^2 + Y^2 + 2\mu^2 I \geq 0$ ، آنگاه Φ_μ در (X, Y, μ) به طور پیوسته مشتق‌پذیر است و

$$\nabla \Phi_\mu(X, Y, \mu)(U, V, \tau) = U + V - \mathcal{L}_C^{-1}[XU + UX + YV + VY + 2\tau\mu I], \quad (5.4)$$

که در آن $C = (X^2 + Y^2 + 2\mu^2 I)^{\frac{1}{2}}$

لم ۴.۲.۴. برای هر $\mu \rightarrow 0_+$ ، $y_* = (x_*^T, z_*^T)^T$ در معادله

$$\eta(y) = \begin{bmatrix} -(\nabla f(x) + \mathcal{A}[\varphi_\mu(a(x), z) + z]) \\ \varphi_\mu(a(x), z) \end{bmatrix} = 0, \quad (6.4)$$

صدق می‌کند اگر و تنها اگر x_* یک نقطه KKT برای مساله نیمه معین (۴.۲) و (۵.۲) باشد.

^۳Continuously differentiable

^۴Generalize the smoothed Fischer-Burmeister function

^۵Strictly monotone

برهان. فرض کنید $\eta(y_*) = 0$ آنگاه

$$\nabla f(x_*) + \mathcal{A}[\varphi_\mu(a(x_*), z_*) + z_*] = 0, \quad (7.4)$$

$$\varphi_\mu(a(x_*), z_*) = 0. \quad (8.4)$$

واضح است که (۸.۴) معادل است با $\Phi_\mu(A(x_*), Z_*) = 0$. حال از لم ۱.۲.۴ به سادگی نتیجه می‌شود

$$Z_* \succeq 0, \quad A(x_*) \succeq 0, \quad \langle Z_*, A(x_*) \rangle = 0, \quad (9.4)$$

زمانی که $\mu \rightarrow 0_+$ از طرف دیگر، با جایگذاری (۷.۴) و (۸.۴) داریم

$$\nabla f(x_*) + \bar{A}z_* = 0. \quad (10.4)$$

توجه کنید $DA(x_*)^* Z_* = (\langle A_1, Z_* \rangle, \dots, \langle A_n, Z_* \rangle)^T = \bar{A}z_*$ بنابراین (۱۰.۴) نتیجه می‌دهد

$$\nabla f(x_*) + DA(x_*)^* Z_* = 0. \quad (11.4)$$

از (۹.۴) و (۱۱.۴)، نتیجه می‌شود که $y_* = (x_*, Z_*)$ در شرایط بهینگی KKT (۲۵.۲) – (۲۷.۲) صدق می‌کند. اثبات عکس قضیه سراسر است. □

حال با استفاده از لم ۴.۲.۴ به سادگی می‌توان نشان داد که برای یافتن جواب سیستم KKT (۲۵.۲) – (۲۷.۲) می‌توانیم مساله مینیم سازی هموار و نامقید زیر را حل کنیم

$$\text{minimize } E(y) = \frac{1}{\rho} \|\eta(y)\|^2 \quad (12.4)$$

لم ۴.۲.۴ نشان می‌دهد که $E(y)$ یک تابع شایسته هموار برای سیستم KKT (۲۵.۲) – (۲۷.۲) بوده و برای هر $y \in \mathbb{R}^{n+m^*}$ به طور پیوسته مشتق پذیر است.

حال فرض کنید $x(\cdot)$ و $Z(\cdot)$ متغیرهایی وابسته به زمان باشند. هدف ما ارائه یک سیستم دینامیکی زمان پیوسته است که مساله یافتن نقاط بهینگی KKT برای مساله نیمه معین و دوگان آن را حل کند که یافتن جواب مساله اصلی متناظر با یافتن نقطه تعادل تابع انرژی $E(y)$ نیز است. بنابراین با استفاده از روش تندترین کاهش برای حل مساله نیمه معین (۴.۲) و (۵.۲) مدل گرادینانی زیر را پیشنهاد می‌دهیم.

$$\frac{dy(t)}{dt} = -\rho \nabla E(y(t)), \quad \rho > 0, \quad (13.4)$$

$$y(0) = y_0, \quad (14.4)$$

که در آن ρ یک پارامتر اسکالر و نشان دهنده نرخ همگرایی شبکه عصبی است. واضح است که هرچه ρ بزرگتر باشد نرخ همگرایی بهتر خواهد بود.

۳.۴ تحلیل پایداری و همگرایی

در این بخش پایداری و همگرایی شبکه عصبی پیشنهاد شده ارائه می‌شود. برای راحتی و بدون اینکه از کلیت مساله کاسته شود فرض می‌کنیم $\rho = 1$.

لم ۱.۳.۴. اگر A یک ماتریس نامنفرد باشد، آن گاه سیستم $AX = \circ$ تنها جواب بدیهی $X = \circ$ دارد.

قضیه ۱.۳.۴. (الف) اگر x_* یک جواب مساله نیمه معین خطی (۴.۲) و (۵.۲) باشد، آن گاه $(x_*^T, z_*^T)^T$ یک نقطه تعادل (۱۳.۴) و (۱۴.۴) است، که در آن Z_* ضربگر لاگرانژ مرتبط با x_* است.

(ب) اگر ماتریس ژاکوبین $\nabla\eta(y)$ از نگاشت η که در (۶.۴) تعریف شده است، نامنفرد باشد و $y_* = (x_*^T, z_*^T)^T$ یک نقطه تعادل (۱۳.۴) و (۱۴.۴) باشد، آن گاه x_* یک جواب مساله نیمه معین خطی (۴.۲) و (۵.۲) است.

برهان. (الف) فرض کنید x_* یک جواب مساله نیمه معین خطی (۴.۲) و (۵.۲) باشد. از لم ۴.۲.۴، واضح است که $\eta(y_*) = \circ$ که $y_* = (x_*^T, z_*^T)^T$ با یک محاسبه ساده داریم

$$\nabla E(y_*) = (\nabla\eta(y_*))^T \eta(y_*) = \circ. \quad (15.4)$$

یعنی y_* یک نقطه تعادل شبکه عصبی (۱۳.۴) و (۱۴.۴) است.

(ب) از لم ۴.۲.۴ و لم ۱.۳.۴ اثبات واضح است.

□

قضیه ۲.۳.۴. نقطه تعادل شبکه عصبی (۱۳.۴) و (۱۴.۴) یکتاست.

برهان. مساله نیمه معین دارای جواب یکتای x_* است. پس سیستم KKT (۲۵.۲) – (۲۷.۲) جواب یکتای y_* دارد. بعلاوه از قضیه ۱.۳.۴ می‌دانیم که جواب سیستم KKT (۲۵.۲) – (۲۷.۲) یک نقطه تعادل شبکه عصبی (۱۳.۴) و (۱۴.۴) است. بنابراین نقطه تعادل شبکه عصبی یکتاست.

□

حال به بیان مهم‌ترین نتایج این فصل می‌پردازیم

قضیه ۳.۳.۴. اگر y_* نقطه تعادل شبکه عصبی (۱۳.۴) و (۱۴.۴) باشد، آنگاه y_* برای (۱۳.۴) و (۱۴.۴) پایدار مجانبی است.

برهان. ابتدا توجه کنید که $E(y) \geq \circ$ و $E(y_*) = \circ$. به‌علاوه، چون y_* یک نقطه تعادل تنهای شبکه عصبی (۱۳.۴) و (۱۴.۴) است، یک همسایگی $\Omega_* \subset \mathbb{R}^{n+m^2}$ از y_* وجود دارد به طوری که

$$\nabla E(y_*) = \circ, \quad \nabla E(y) \neq \circ, \quad \forall y \in \Omega_* \setminus \{y_*\}.$$

در نتیجه برای هر $y \in \Omega_* \setminus \{y_*\}$ ، $E(y) > 0$. در غیر این صورت اگر برای یک $y \in \Omega_* \setminus \{y_*\}$ داشته باشیم $\nabla E(y) = 0$ یعنی y نیز یک نقطه تعادل شبکه عصبی (۱۳.۴) و (۱۴.۴) است که با فرض تنها بودن نقطه تعادل y_* در Ω در تناقض است. علاوه بر آن

$$\frac{dE(y(t))}{dt} = [\nabla E(y(t))]^T \frac{dy(t)}{dt} = -\rho \|\nabla E(y(t))\|^2 \leq 0, \quad (16.4)$$

و

$$\frac{dE(y(t))}{dt} < 0, \forall y(t) \in \Omega, \quad y(t) \neq y_*. \quad (17.4)$$

با بکاربردن قضیه ۸.۳.۱ نتیجه می‌گیریم که y_* پایدار مجانبی است.

□

قضیه ۴.۳.۴. فرض کنید $y = y(t, y_0)$ یک مسیر شبکه عصبی (۱۳.۴) و (۱۴.۴) است که $y_0 = y(0, y_0)$ و مجموعه سطح $L(y_0) = \{y \in \mathbb{R}^{n+m^*} : E(y) \leq E(y_0)\}$ کراندار است، آنگاه

(الف) $\gamma^+ = \{y(t, y_0) \mid t \geq 0\}$ کراندار است.

(ب) وجود دارد \bar{y} به طوری که $\lim_{t \rightarrow \infty} y(t, y_0) = \bar{y}$.

برهان. (الف) فرض کنید y_* یک نقطه تعادل شبکه عصبی (۱۳.۴) و (۱۴.۴) باشد. مشتق $E(y)$ را در طول مسیر $y(t, y_0)$ ، $(t \geq 0)$ محاسبه می‌کنیم. از (۱۰.۴)، $E(y)$ در طول مسیر $y(t, y_0)$ ، $(t \geq 0)$ به طور یکنواخت ناصعودی است. بنابراین $L(y_0) \subset \gamma^+(y_0)$. این بیان می‌کند که $\gamma^+ = \{y(t, y_0) \mid t \geq 0\}$ کراندار است.

(ب) از (الف) $\gamma^+ = \{y(t, y_0) \mid t \geq 0\}$ یک مجموعه کراندار است. دنباله صعودی یکنواخت $\{\bar{t}_n\}$ ، $\bar{t}_1 \leq \dots \leq \bar{t}_n \leq \dots$ و $\bar{t}_n \rightarrow +\infty$ را در نظر می‌گیریم. آنگاه $\{y(\bar{t}_n, y_0)\}$ دنباله‌ای کراندار از بی‌نهایت نقطه است. بنابراین یک نقطه حدی \bar{y} وجود دارد، یعنی اینکه یک زیر دنباله مانند $\{t_{nk}\} \subset \{\bar{t}_n\}$ ، $t_{nk} \rightarrow +\infty$ وجود دارد به طوری که $\lim_{k \rightarrow +\infty} y(t_{nk}, y_0) = \bar{y}$ که در آن \bar{y} در $\frac{dE(y(t))}{dt} = 0$ صدق می‌کند این نشان می‌دهد \bar{y} یک $-\omega$ نقطه حدی از $\gamma^+(y_0)$ است. با بکار بردن اصل تغییرناپذیری لسال در ۱۰.۳.۱ داریم آنگاه $y(t, y_0) \rightarrow \bar{y} \in M$ وقتی که $t \rightarrow \infty$ که در آن M بزرگترین مجموعه تغییرناپذیر در $K = \{y(t, y_0) \mid \frac{dE(y(t, y_0))}{dt} = 0\}$ است. از (۱۳.۴)، (۱۴.۴) و (۱۶.۴) نتیجه می‌شود که

$$\frac{dE(y(t))}{dt} = 0 \iff \frac{dx}{dt} = 0, \quad \frac{dz}{dt} = 0. \quad (18.4)$$

بنابراین $\bar{y} \rightarrow D^*$ که $M \subset K \subset D^*$. در نتیجه با هر نقطه شروع y_0 ، منحنی $y(t, y_0)$ شبکه عصبی به \bar{y} میل می‌کند. این اثبات را کامل می‌کند.

□

به عنوان نتیجه‌ای از دو قضیه‌ی اخیر داریم

نتیجه ۱.۳.۴. اگر $D^* = \{(x^*, Z^*)\}$ ، آنگاه شبکه عصبی (۱۳.۴) و (۱۴.۴) پایدار مجانبی سراسری به نقطه تعادل منحصر به فرد $y^* = (x^*, z^*)$ است که در آن نشان دهنده مجموعه نقاط بهینه مساله نیمه معین (۴.۲) و (۵.۲) است.

۴.۴ مثال های عددی

در این بخش مثال هایی ارائه خواهد شد تا نتایجی را که در بخش های قبل حاصل شد نشان دهند. برای این مسائل، ما همچنین کارایی عددی شبکه عصبی (۱۳.۴) و (۱۴.۴) را با مقادیر مختلف ρ و نقاط آغازین مختلف y_0 مقایسه می کنیم. شبیه سازی این مثال ها با استفاده از نرم افزار Matlab 7، و حل کننده معادلات دیفرانسیل معمولی ode45 انجام شده است.

مثال ۲.۴ [۸] مساله برنامه ریزی خطی زیر را در نظر بگیرید

$$\begin{aligned} & \text{minimize} && 3x_1 - x_2 \\ & \text{subject to} && x_1 - 3x_2 \geq -3, \\ & && 2x_1 + x_2 \geq -2, \\ & && 2x_1 + x_2 \leq 8, \\ & && 4x_1 - x_2 \leq 16. \end{aligned}$$

جواب بهینه این مساله $x^* = (0/7143, -1/3571)^T$ است. همان طور که در فصل قبل دیدیم این مساله می تواند به عنوان مسئله نیمه معین (۴.۲) و (۵.۲) طرح شود که

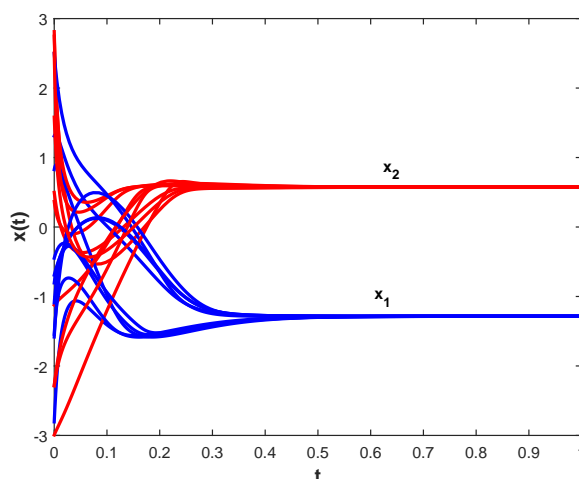
$$A_0 = \text{diag}(3, 2, 8, 16), \quad A_1 = \text{diag}(-1, -2, 2, 4), \quad A_2 = \text{diag}(3, -1, 1, -1).$$

شبکه عصبی (۱۳.۴) و (۱۴.۴) را برای حل این مسئله به کار برده ایم. شکل ۱.۴ رفتار همگرایی مسیرهای گذرای متغیر وضعیت $x_1(t)$ و $x_2(t)$ را نشان می دهد.

مثال ۲.۴. مساله مینیمم کردن بزرگترین مقدار ویژه ماتریس $A(x) = A_0 + \sum_{i=1}^5 x_i A_i$ در مثال ۳.۳ را در نظر بگیرید. همانطور که در فصل ۳ دیدیم این مساله می تواند به عنوان یک مسئله نیمه معین به صورت زیر طرح شود

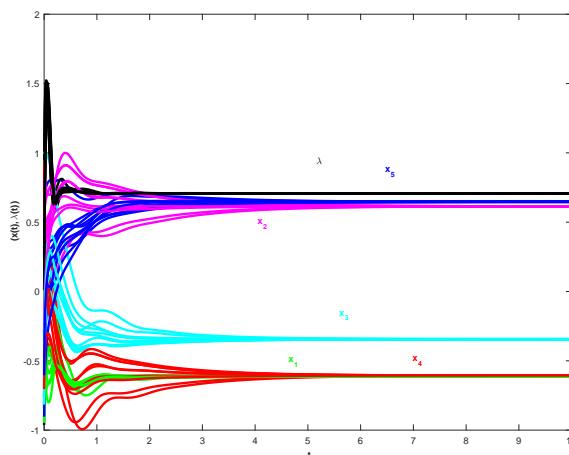
$$\begin{aligned} & \text{minimize} && \lambda \\ & \text{subject to} && B(\lambda, x) = \lambda I - A(x) \succeq 0. \end{aligned}$$

جواب بهینه این مساله $x^* = (-0/6136, 0/6145, -0/3437, -0/6067, 0/6465)^T$ و مقدار بهینه آن $\lambda^* = 0/7089$ است. شکل ۲.۴ نشان می دهد که مسیرهای شبکه عصبی (۱۳.۴) و (۱۴.۴) برای حل این مساله با 20 نقطه شروع تصادفی به جواب بهینه مساله همگراست. در این مساله تاثیر پارامتر ρ روی $\|x(t) - x^*\|^2$ را در شبکه عصبی (۱۳.۴) و (۱۴.۴) بررسی



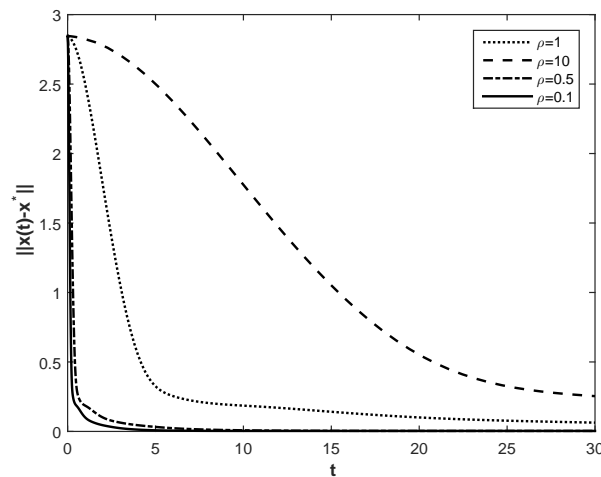
شکل ۱.۴: رفتار همگرایی $(x_1(t), x_2(t))^T$ از شبکه عصبی (۱۳.۴) و (۱۴.۴) با $\rho = 10$ نقطه آغازین تصادفی در مثال ۱.۴

می‌کنیم. شکل ۳.۴ رفتار نرم خطای $\|x(t) - x^*\|^2$ را با ۴ مقدار مختلف برای ρ و نقطه شروع عصبی (۱۴.۴) و (۱۳.۴) همگرایی کندتری نسبت به زمانی که $\rho = 10$ است دارد. همچنین شکل ۴.۴ رفتار نرم خطای $\|x(t) - x^*\|^2$ را با $\rho = 20$ نقطه شروع مختلف نشان می‌دهد.



شکل ۲.۴: رفتار همگرایی $(\lambda(t), x(t)^T)^T$ از شبکه عصبی (۱۳.۴) و (۱۴.۴) با $\rho = 20$ نقطه آغازین تصادفی در مثال ۲.۴

مثال ۳.۴. ماتریس‌های A_i ، $i = 0, 1, \dots, 5$ ، در مثال ۳.۳ را در نظر بگیرید. مسئله مینیمم کردن فاصله بین بزرگترین و کوچکترین مقدار ویژه ماتریس $A(x)$ است. در فصل قبل دیدیم



شکل ۳.۴: رفتار همگرایی $\|(\lambda(t), x(t)) - (\lambda^*, x^*)\|^2$ با $(\lambda_0, x_0^T)^T = (2, -1, 1, -1, 0, 1)^T$ در مثال ۲.۴.

که این مساله می تواند به صورت مساله نیمه معین زیر طرح شود

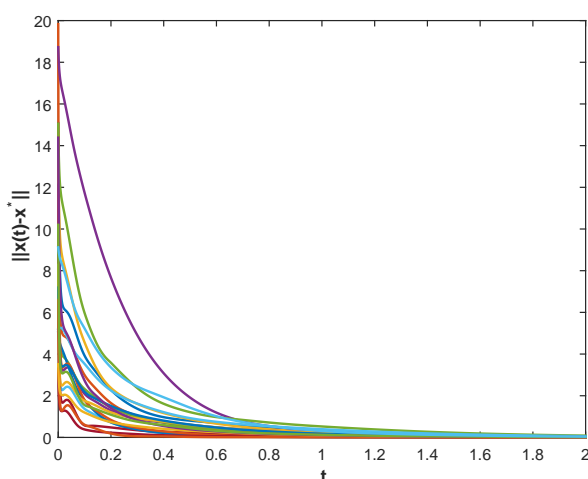
$$\begin{aligned} & \text{minimize } \lambda - \mu \\ & \text{subject to } \mu I \preceq A(x) \preceq \lambda I, \end{aligned}$$

که $\lambda = \lambda_{max}(A(x))$ و $\mu = \lambda_{min}(A(x))$ به ترتیب بزرگترین و کوچکترین مقدار ویژه ماتریس $A(x)$ هستند. تمام نتایج عددی نشان می دهند که شبکه عصبی (۱۳.۴) و (۱۴.۴) به جواب بهینه $(\lambda^*, \mu^*, x^{*T})^T$ همگراست و مقدار بهینه تابع هدف $1/5389$ است. شکل ۵.۴ رفتار همگرایی مسیرهای گذرای $(\lambda(t), \mu(t), x(t)^T)^T$ و تابع هدف را با 20 نقطه آغازین تصادفی نشان می دهد. همگرایی سریع به جواب بهینه به خوبی مشاهده می شود. شکل ۶.۴ تاثیر پارامتر ρ را بر مقدار $\|x(t) - x^*\|^2$ نشان می دهد. از شکل ۶.۴ ما می بینیم زمانی که $\rho = 0.05$ شبکه عصبی (۱۳.۴) و (۱۴.۴) همگرایی کندتری از $\|x(t) - x^*\|^2$ را نسبت به زمانی که $\rho = 20$ نتیجه می دهد.

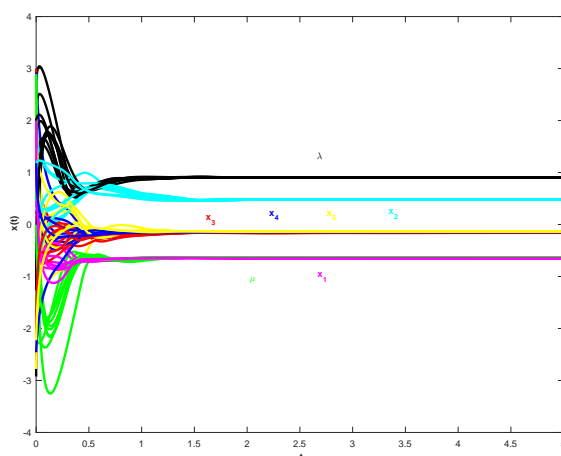
مثال ۴.۴. مساله ما مینیمم کردن $\|A(x)\|$ ، نرم ماتریس $A(x) = A_0 + x_1 A_1 + \dots + x_5 A_5$ ، است که A_i ، $i = 0, 1, \dots, 5$ در مثال ۳.۳ داده شدند. در فصل قبل گفتیم که این مساله می تواند با مساله نیمه معین زیر طرح شود

$$\begin{aligned} & \text{minimize } \nu \\ & \text{subject to } B(\nu, x) = \begin{pmatrix} \nu I & A(x) \\ A(x)^T & \nu I \end{pmatrix} \succeq 0. \end{aligned}$$

ما شبکه عصبی (۱۳.۴) و (۱۴.۴) را برای حل این مساله به کار می گیریم. تمام نتایج نشان می دهند که شبکه عصبی (۱۳.۴) و (۱۴.۴) به جواب بهینه $(\nu^*, x^{*T})^T$ همگراست و

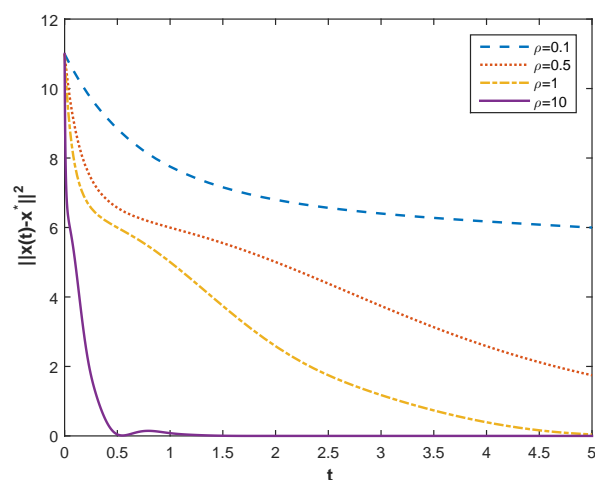


شکل ۴.۴: رفتار همگرایی $(\lambda, \mu, x^T)^T$ شبکه عصبی (۱۳.۴) و (۱۴.۴) با ν_0 نقطه آغازین تصادفی در مثال ۳.۴.

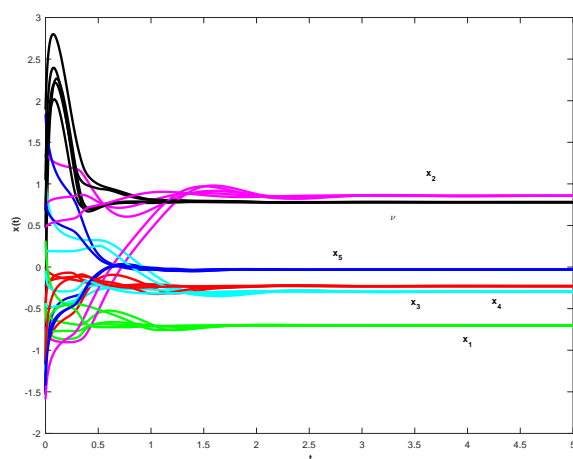


شکل ۵.۴: رفتار همگرایی تابع هدف با ν_0 نقطه آغازین تصادفی در مثال ۳.۴.

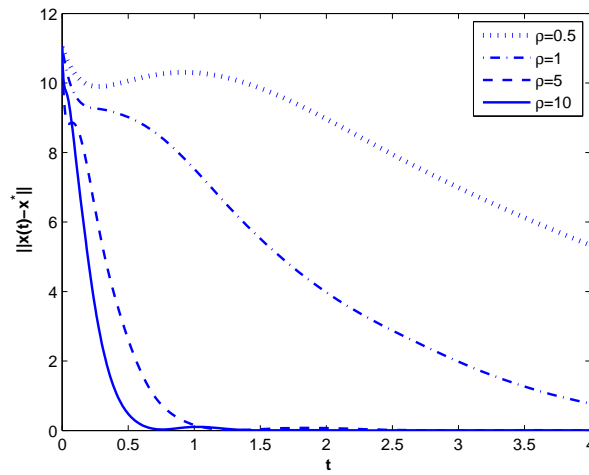
مقدار بهینه تابع هدف 0.7806 است. برای مثال، شکل ۷.۴ منحنی‌های متغیرهای وضعیت مقدار بهینه تابع هدف $(\nu, x(t)^T)^T$ را با ν_0 نقطه آغازین تصادفی نشان می‌دهد. همگرایی سریع شبکه عصبی مشاهده می‌شود. شکل ۸.۴ رفتار همگرایی منحنی خطا را با مقادیر مختلف ρ و با نقطه آغازین $(\nu_0, x_0^T)^T = (2, 1, 2, -2, 1, 1)^T$ نشان می‌دهد. واضح است که ρ بزرگتر، نرخ همگرایی بهتر را نتیجه می‌دهد.



شکل ۶.۴: رفتار همگرایی $\|x(t) - x^*\|^2$ با $(\lambda_0, \mu_0, x_0^T)^T = (1, -1, 1, -1, 0, 2, 1)^T$ در مثال ۳.۴.



شکل ۷.۴: رفتار همگرایی $(\nu(t), x(t)^T)^T$ در شبکه عصبی (۱۳.۴) و (۱۴.۴) با ۵ نقطه آغازین تصادفی در مثال ۴.۴.



شکل ۸.۴: رفتار همگرایی $\|(\nu(t), x(t)^T)^T - (\nu^*, x^{*T})^T\|^2$ با $(\nu_0, x_0^T)^T$ در مثال ۴.۴ $(2, 1, 2, -2, 1, 1)^T$.

فصل ۵

یک مدل شبکه عصبی برای حل مساله نیمه معین غیرخطی

۱.۵ مقدمه

در این فصل مساله برنامه‌ریزی نیمه معین غیرخطی^۱ را در نظر می‌گیریم

$$\text{minimize } f(x) \quad (1.5)$$

$$\text{subject to } A(x) = A_0 + \sum_{i=1}^m x_i A_i \preceq 0, \quad (2.5)$$

$$g(x) \leq 0, \quad (3.5)$$

که در آن $f: \mathbb{R}^n \rightarrow \mathbb{R}$ و $g: \mathbb{R}^n \rightarrow \mathbb{R}^q$ به طور پیوسته مشتق پذیر هستند و $A_i \in \mathbb{R}^{m \times m}$ ، $i = 1, 2, \dots, n$. توجه داشته باشید که مسئله نیمه معین غیرخطی (۱.۵)–(۳.۵) یک مسئله بهینه‌سازی محدب است.

مسائل نیمه معین غیرخطی اغلب در شاخه‌های کاربردی گسترده‌ای ظاهر می‌شوند مانند

^۱Nonlinear Semidefinite Programming(NSDP)

طراحی ساختاری بهینه^۲، کنترل بهینه مقاوم^۳، طراحی کنترل بازخورد مقاوم^۴، مسائل طراحی اطمینان^۵ و بهینه سازی مقدار ویژه^۶ [۱۱]، [۲۶] و [۲۷] را ببینید). بنابراین مسائل نیمه معین غیرخطی در دهه‌های اخیر بسیار مورد توجه قرار گرفته‌اند. تاکنون روش‌های مختلفی برای حل این مسائل به کار رفته‌اند، مانند روش ضرایب جریمه‌ای مانعی^۷، روش برنامه‌ریزی نیمه معین متوالی^۸ روش لاگرانژ افزوده شده^۹، و روش نقطه درونی اولیه^{۱۰} (می‌توانید مقالات [۱۷]، [۹۸] و [۹۹] را ببینید).

در فصل‌های قبل گفتیم که یک رهیافت برای حل مسائل بهینه سازی استفاده از شبکه‌های عصبی مصنوعی است. در مقایسه با الگوریتم‌های عددی، شبکه‌های عصبی مصنوعی دارای قابلیت انجام محاسبات موازی و همگرایی سریع هستند. هر چند انواع مختلف شبکه‌های عصبی برای حل مسائل مختلف پیشنهاد شده‌اند، اما تاکنون هیچ گزارشی از به کارگیری شبکه‌های عصبی برای حل مسائل نیمه معین غیرخطی ارائه نشده است. بنابراین با توجه به مزایای شبکه‌های عصبی ساختن یک شبکه عصبی برای حل مسئله نیمه معین غیرخطی (۱.۵)–(۳.۵) ضروری است. ایده اصلی در این فصل ارائه یک سیستم دینامیکی برای حل مسئله نیمه معین غیرخطی (۱.۵)–(۳.۵) است. به این منظور بر اساس شرایط بهینگی KKT مسئله نیمه معین غیرخطی (۱.۵)–(۳.۵) یک مدل شبکه عصبی برای حل مسئله نیمه معین غیرخطی (۱.۵)–(۳.۵) پیشنهاد می‌کنیم. پایداری لیاپانوف و همگرایی سراسری مدل شبکه عصبی پیشنهادی را اثبات کرده و مثال‌هایی عددی برای نشان دادن کارایی مدل پیشنهادی ارائه می‌دهیم.

۲.۵ مدل شبکه عصبی

در این بخش با استفاده از یک روش استاندارد بهینه سازی، مساله برنامه‌ریزی نیمه معین غیرخطی (۱.۵)–(۳.۵) را به یک سیستم دینامیکی غیرخطی تبدیل می‌کنیم. مسئله نیمه معین غیرخطی (۱.۵)–(۳.۵) را به عنوان مسئله اولیه در نظر می‌گیریم. شرایط لازم بهینگی KKT برای یک نقطه شدنی از مسئله نیمه معین غیرخطی (۱.۵)–(۳.۵) عبارتست از وجود یک $u_* \in \mathbb{R}^q$ و $Z_* \in S^m$ است به طوری که

^۲Optimal structural design
^۳Optimal robust control
^۴Robust feedback control design
^۵Trust design problems
^۶Eigenvalue optimization
^۷Penalty/barrier multiplier method
^۸Sequential SDP method
^۹Augmented Lagrangian method
^{۱۰}Primal-dual interior point method

$$\nabla f(x_*) + DA(x_*)^* Z_* + \nabla g(x_*)^T u_* = \circ, \quad (۴.۵)$$

$$u_*^T g(x_*) = \circ, \quad u_* \geq \circ, \quad (۵.۵)$$

$$\langle Z_*, A(x_*) \rangle = \circ, \quad Z_* \succeq \circ. \quad (۶.۵)$$

x_* نقطه KKT مسئله نیمه معین غیرخطی (۱.۵) - (۳.۵) و زوج (u_*, Z_*) ضربگر لاگرانژ متناظر با x_* نامیده می‌شود.

همانند مسئله نیمه معین خطی در این فصل نیز فرض می‌کنیم سه شرط زیر برقرارند

۱. ماتریس‌های $A_i, i = 1, \dots, n$ ، مستقل خطی هستند، یعنی $\sum_{i=1}^n u_i A_i = \circ$ نتیجه می‌دهد $i = 1, \dots, n, u_i = \circ$.

۲. هر دو مسئله اولیه و دوگان شدنی هستند، یعنی

$$\exists Z_o \in S^m, \quad x_o \in \mathbb{R}^n : Z_o \succ \circ, \quad A(x_o) \prec \circ, \quad \langle -A_i, Z_o \rangle = c_i, i = 1, \dots, n.$$

۳. مجموعه گرادیان‌های $\{\nabla g_k(x) | k = 1, \dots, q\}$ مستقل خطی هستند.

اکنون متغیرهای وابسته به زمان $x(\cdot), Z(\cdot)$ و $u(\cdot)$ را در نظر بگیرید. ما یک مدل شبکه عصبی بازگشتی برای حل مساله برنامه‌ریزی نیمه معین غیرخطی (۱.۵) - (۳.۵) ارائه می‌دهیم که معادله دینامیکی آن به صورت زیر داده شده است:

$$\frac{dx}{dt} = -(\nabla f(x) + DA(x)^*[Z + A(x)]_+ + \nabla g(x)^T(u + g(x))^+), \quad (۷.۵)$$

$$\frac{dZ}{dt} = [Z + A(x)]_+ - Z, \quad (۸.۵)$$

$$\frac{du}{dt} = (u + g(x))^+ - u, \quad (۹.۵)$$

که $DA(x) = [A_1, A_2, \dots, A_n]$ و

$$(u + g(x))^+ = ((u_1 + g_1(x))^+, (u_2 + g_2(x))^+, \dots, (u_q + g_q(x))^+),$$

$$(u_k + g_k(x))^+ = \max(u_k + g_k(x), \circ), \quad k = 1, 2, \dots, q.$$

فرض کنید $z = \text{vec}(Z)$ ، $a^i = \text{vec}(A_i)$ ، $(i = 1, \dots, n)$ ، $a(x) = \text{vec}(A(x))$ ، $(z + a(x))_+ = \text{vec}([Z + A(x)]_+)$ و $\bar{A} = [a^1 \ a^2 \ \dots \ a^n]^T$. با استفاده از (۲.۲) و (۳.۲)، شبکه عصبی (۷.۵) - (۹.۵) به صورت زیر نوشته می‌شود

$$\frac{dx}{dt} = -(\nabla f(x) + \bar{A}(z + a(x))_+ + \nabla g(x)^T(u + g(x))^+),$$

$$\frac{dz}{dt} = (z + a(x))_+ - z,$$

$$\frac{du}{dt} = (u + g(x))^+ - u.$$

برای ساده شدن فرض کنید $y = (x^T, z^T, u^T)^T \in \mathbb{R}^{n+m^2+q}$ و

$$\Phi(y) = \begin{bmatrix} -(\nabla f(x) + \bar{A}(z + a(x))_+ + \nabla g(x)^T(u + g(x))^+) \\ (z + a(x))_+ - z \\ (u + g(x))^+ - u \end{bmatrix} \quad (10.5)$$

بنابراین شبکه عصبی (۷.۵) - (۹.۵) به صورت زیر نوشته می شود

$$\frac{dy}{dt} = \rho \Phi(y), \quad (11.5)$$

$$y(t_0) = y_0, \quad \rho > 0, \quad (12.5)$$

که $y_0 \in S_0 = \{y = (x^T, z^T, u^T)^T \in \mathbb{R}^{n+m^2+q} | Z \in S_+^m \text{ \& } z = \text{vec}(Z)\}$ و ρ یک پارامتر اسکالر و نشان دهنده نرخ همگرایی شبکه عصبی (۱۱.۵) و (۱۲.۵) است. واضح است که هرچه ρ بزرگتر باشد نرخ همگرایی بهتر خواهد بود.

۳.۵ تحلیل پایداری و همگرایی

در این بخش پایداری و همگرایی شبکه عصبی پیشنهاد شده (۱۱.۵) و (۱۲.۵) ارائه می شود. برای راحتی و بدون اینکه از کلیت مساله کاسته شود فرض می کنیم $\rho = 1$ باشد.

قضیه ۱.۳.۵. فرض کنید $y^* = (x^{*T}, z^{*T}, u^{*T})^T$ یک نقطه تعادل شبکه عصبی (۱۱.۵) و (۱۲.۵) باشد، آنگاه x^* یک نقطه KKT از مسئله نیمه معین غیرخطی (۱.۵) - (۳.۵) است. از طرف دیگر، اگر $x^* \in \mathbb{R}^n$ یک جواب بهینه از مسئله نیمه معین غیرخطی (۱.۵) - (۳.۵) باشد، آنگاه یک $Z^* \in S_+^m$ و $u^* \in \mathbb{R}^q$ وجود دارد به طوری که $y^* = (x^{*T}, z^{*T}, u^{*T})^T$ یک نقطه تعادل شبکه عصبی (۱۱.۵) و (۱۲.۵) است که $z^* = \text{vec}(Z^*)$.

برهان. فرض کنید $y^* = (x^{*T}, z^{*T}, u^{*T})^T$ یک نقطه تعادل شبکه عصبی (۱۱.۵) و (۱۲.۵) باشد. آنگاه $\frac{dx^*}{dt} = 0$ ، $\frac{dz^*}{dt} = 0$ و $\frac{du^*}{dt} = 0$ به راحتی نتیجه می شود

$$\nabla f(x^*) + \bar{A}(z^* + a(x^*))_+ + \nabla g(x^*)^T(u^* + g(x^*))_+ = 0 \quad (13.5)$$

$$(z^* + a(x^*))_+ = z^*, \quad (14.5)$$

$$(u^* + g(x^*))_+ = u^*. \quad (15.5)$$

واضح است که $(u^* + g(x^*))_+ = u^*$ اگر و فقط اگر

$$u^* \geq 0, \quad g(x^*) \leq 0, \quad u^{*T}g(x^*) = 0. \quad (16.5)$$

همچنین $(z^* + a(x^*))_+ = z^*$ معادل است با $[Z^* + A(x^*)]_+ = Z^*$. بنابراین از لم (۲.۲.۲) داریم

$$Z^* \succeq 0, \quad A(x^*) \preceq 0, \quad \langle Z^*, A(x^*) \rangle = 0. \quad (17.5)$$

با (۳.۲) داریم $DA(x^*)^*Z^* = \bar{A}z^*$ و با جایگذاری (۱۴.۵) و (۱۵.۵) در (۱۳.۵) داریم

$$\nabla f(x^*) + DA(x^*)^*Z^* + \nabla g(x^*)^T u^* = 0. \quad (18.5)$$

از (۱۶.۵) – (۱۸.۵)، می‌بینیم که $y^* = (x^{*T}, z^{*T}, u^{*T})^T$ در شرایط بهینگی (۴.۵)–(۶.۵) صدق می‌کند. اثبات عکس قضیه سراسر است. \square

قضیه ۲.۳.۵. برای هر نقطه شروع $y_0 = y(t_0) \in S_0$ ، یک جواب پیوسته برای سیستم (۱۱.۵) و (۱۲.۵) وجود دارد.

برهان. چون $\nabla f(x)$ ، $DA(x)$ و $\nabla g(x)$ روی مجموعه محدب باز $B \subseteq \mathbb{R}^{n+m^2+q}$ پیوسته هستند $(u + g(x))^+ - u$ و $(z + a(x))_+ - z$ ، $\nabla f(x) + \bar{A}(z + a(x))_+ + \nabla g(x)^T(u + g(x))^+$ لیب شیتز محلی هستند. طبق قضیه وجود محلی معادلات دیفرانسیل معمولی ۵.۳.۱، شبکه عصبی (۱۱.۵) و (۱۲.۵) یک جواب پیوسته منحصر به فرد $y(t)$ ، برای $t \in [t_0, \tau)$ ، $\tau > t_0$ دارد. \square

قضیه ۳.۳.۵. جواب شبکه عصبی (۱۱.۵) و (۱۲.۵) با نقطه شروع $y_0 \in S_0$ پایدار لیپانوف است. علاوه بر این، مسیر جواب شبکه عصبی (۱۱.۵) و (۱۲.۵) جواب ماکزیمال نیست.

برهان. قرار دهید $\bar{z} = (z + a(x))_+$ و $\bar{x} = x - \nabla f(x) - \bar{A}\bar{z}$ ، $\bar{u} = (u + g(x))^+$ و تابع لیپانوف زیر را در نظر بگیرید

$$E(y) = f(x) - f(x^*) + \frac{1}{\rho} [\|\bar{z}\|^2 - \|z^*\|^2] + \frac{1}{\rho} [\|\bar{u}\|^2 - \|u^*\|^2] - (x - x^*)^T (\nabla f(x^*) + \bar{A}z^* + \nabla g(x^*)u^*) - (z - z^*)^T z^* - (u - u^*)^T u^* + \frac{1}{\rho} \|x - x^*\|^2 + \frac{1}{\rho} \|z - z^*\|^2 + \frac{1}{\rho} \|u - u^*\|^2 \quad (19.5)$$

که $y^* = (x^{*T}, z^{*T}, u^{*T})^T$ یک نقطه تعادل شبکه عصبی (۱۱.۵) و (۱۲.۵) باشد. آنگاه

$$\begin{aligned}
 \frac{dE(y(t))}{dt} &= [\nabla f(x) + \bar{A}\bar{z} + \nabla g(x)^T \bar{u} - \nabla f(x^*) - \bar{A}z^* - \nabla g(x^*)^T u^* + x - x^*]^T (\bar{x} - x) \\
 &+ (\bar{z} + z - \Upsilon z^*)^T (\bar{z} - z) / \Upsilon + (\bar{u} + u - \Upsilon u^*)^T (\bar{u} - u) / \Upsilon \\
 &= -\|x - \bar{x}\|^{\Upsilon} + [\nabla f(x) + \bar{A}\bar{z} + \nabla g(x)^T \bar{u} - \nabla f(x^*) - \bar{A}z^* - \nabla g(x^*)^T u^* + \bar{x} - x^*]^T (\bar{x} - x) \\
 &- \frac{1}{\Upsilon} \|z - \bar{z}\|^{\Upsilon} + (\bar{z} - z^*)^T (\bar{z} - z) + (\bar{u} - u^*)^T (\bar{u} - u) \\
 &= -\|x - \bar{x}\|^{\Upsilon} - [\nabla f(x) - \nabla f(x^*)]^T (x - x^*) - (\nabla f(x^*) + \bar{A}z^* + \nabla g(x^*)^T u^*)^T (\bar{x} - x^*) \\
 &- [x - \nabla f(x) - \bar{A}\bar{z} - \nabla g(x)^T \bar{u} - \bar{x}]^T (\bar{x} - x^*) - (\bar{A}\bar{z} - \bar{A}z^*)^T (x - x^*) \\
 &- (\nabla g(x)^T \bar{u} - \nabla g(x^*)^T u^*)^T (x - x^*) \\
 &- \frac{1}{\Upsilon} \|z - \bar{z}\|^{\Upsilon} + (\bar{z} - z^*)^T (a(x) - (z + a(x))_-) - \frac{1}{\Upsilon} \|u - \bar{u}\|^{\Upsilon} + (\bar{u} - u^*)^T (g(x) - (u + g(x))_-) \\
 &= -\|x - \bar{x}\|^{\Upsilon} - \frac{1}{\Upsilon} \|z - \bar{z}\|^{\Upsilon} - \frac{1}{\Upsilon} \|u - \bar{u}\|^{\Upsilon} - [\nabla f(x) - \nabla f(x^*)]^T (x - x^*) \\
 &- (\nabla f(x^*) + \bar{A}z^* + \nabla g(x^*)^T u^*)^T (\bar{x} - x^*) - [x - \nabla f(x) - \bar{A}\bar{z} - \nabla g(x)^T \bar{u} - \bar{x}]^T (\bar{x} - x^*) \\
 &- \bar{z}^T [-\bar{A}^T(x^* - x) - a(x) + a(x^*)] + \bar{z}^T a(x^*) - \bar{z}(z + a(x))_- \\
 &+ z^{*T} [\bar{A}^T(x - x^*) - a(x) + a(x^*)] - z^* a(x^*) + z^*(z + a(x))_- \\
 &- \bar{u}^T [-\nabla g(x)^T (x^* - x) - g(x) + g(x^*)] + \bar{u}^T g(x^*) - \bar{u}(u + g(x))_- \\
 &+ u^{*T} [\nabla g(x)^T (x - x^*) - g(x) + g(x^*)] - u^* g(x^*) + u^*(u + g(x))_- \\
 &= -\|x - \bar{x}\|^{\Upsilon} - \frac{1}{\Upsilon} \|z - \bar{z}\|^{\Upsilon} - \frac{1}{\Upsilon} \|u - \bar{u}\|^{\Upsilon} - [\nabla f(x) - \nabla f(x^*)]^T (x - x^*) \\
 &- (\nabla f(x^*) + \bar{A}z^* + \nabla g(x^*)^T u^*)^T (\bar{x} - x^*) - [x - \nabla f(x) - \bar{A}\bar{z} - \nabla g(x)^T \bar{u} - \bar{x}]^T (\bar{x} - x^*) \\
 &- \bar{z}^T [-\bar{A}^T(x^* - x) - a(x) + a(x^*)] + \bar{z}^T a(x^*) \\
 &- z^{*T} [-\bar{A}^T(x - x^*) + a(x) - a(x^*)] + z^*(z + a(x))_- \\
 &- \bar{u}^T [-\nabla g(x^*)^T (x^* - x) - g(x) + g(x^*)] + \bar{u}^T g(x^*) \\
 &- u^{*T} [-\nabla g(x^*)^T (x - x^*) + g(x) - g(x^*)] + u^*(u + g(x))_-
 \end{aligned}$$

توجه داشته باشید که $\bar{z} - z = a(x) + (-z - a(x))_+$ و $\bar{u} - u = g(x) + (-u - g(x))_+$ بعلاوه به راحتی می توان نشان داد

$$\begin{aligned}
 \bar{z}^T (-z - a(x))_+ &= \circ, z^{*T} a(x^*) = \circ, \\
 z^{*T} (-z - a(x))_+ &\geq \circ, \bar{z}^T a(x^*) \leq \circ.
 \end{aligned}$$

$$\begin{aligned}
 \bar{u}^T (-u - g(x))_+ &= \circ, u^{*T} g(x^*) = \circ, \\
 u^{*T} (-u - g(x))_+ &\geq \circ, \bar{u}^T g(x^*) \leq \circ.
 \end{aligned}$$

پس داریم

$$\begin{aligned} \frac{dE(y(t))}{dt} &\leq -\|x - \bar{x}\|^2 - \frac{1}{\rho}\|z - \bar{z}\|^2 - \frac{1}{\rho}\|u - \bar{u}\|^2 - [\nabla f(x) - \nabla f(x^*)]^T(x - x^*) \\ &- (\nabla f(x^*) + \bar{A}z^* + \nabla g(x^*)^T u^*)^T(\bar{x} - x^*) - [x - \nabla f(x) - \bar{A}\bar{z} - \nabla g(x)^T \bar{u} - \bar{x}]^T(\bar{x} - x^*) \\ &- \bar{z}^T[-\bar{A}^T(x^* - x) - a(x) + a(x^*)] - z^{*T}[-\bar{A}^T(x - x^*) + a(x) - a(x^*)] \\ &- \bar{u}^T[-\nabla g(x)^T(x^* - x) - g(x) + g(x^*)] \\ &- u^{*T}[-\nabla g(x^*)^T(x - x^*) + g(x) - g(x^*)] \end{aligned} \quad (20.5)$$

قرار دهید $u = x^*$ و $v = x - \nabla f(x) - \bar{A}\bar{z} - \nabla g(x)^T \bar{u}$ ، آنگاه با $\Omega = \mathbb{R}^n$ ، با استفاده از لم ۲ در [۱۰۹] داریم

$$(x - \nabla f(x) - \bar{A}\bar{z} - \nabla g(x)^T \bar{u} - \bar{x})^T(\bar{x} - x^*) \geq 0. \quad (21.5)$$

از محدب و مشتق پذیر بودن $f(x)$ و $g(x)$ ، برای هر $x \in \Omega$ داریم

$$\left\{ \begin{array}{l} [\nabla f(x) - \nabla f(x^*)]^T(x - x^*) \geq 0, \\ g(x^*) - g(x) - \nabla g(x)^T(x^* - x) \geq 0, \\ g(x) - g(x^*) - \nabla g(x^*)^T(x - x^*) \geq 0. \end{array} \right. \quad (22.5)$$

همچنین از محدب و مشتق پذیر بودن $A(x)$ داریم

$$\left\{ \begin{array}{l} A(x^*) \leq A(x) + DA(x)^*(x^* - x), \\ A(x) \leq A(x^*) + DA(x^*)^*(x - x^*). \end{array} \right. \quad (23.5)$$

از آنجایی که $\bar{Z}, Z^* \in S_+^m$ ، با استفاده از لم (۲.۲.۲) داریم

$$\left\{ \begin{array}{l} \bar{z}^T[a(x^*) - a(x) + \bar{A}^T(x^* - x)] \leq 0, \\ z^{*T}[a(x) - a(x^*) - \bar{A}^T(x - x^*)] \leq 0. \end{array} \right. \quad (24.5)$$

با جایگذاری (۲۱.۵) - (۲۴.۵) و (۴.۵) در (۲۰.۵) نتیجه می‌گیریم

$$\frac{dE(y(t))}{dt} \leq -\|x - \bar{x}\|^2 - \frac{1}{\rho}\|z - \bar{z}\|^2 - \frac{1}{\rho}\|u - \bar{u}\|^2 \leq 0. \quad (25.5)$$

این یعنی شبکه عصبی (۱۱.۵) و (۱۲.۵) پایدار لیپانوف است. علاوه بر این

$$E(y_0) \geq E(y(t)) \geq \frac{1}{\rho}\|y(t) - y^*\|^2, \quad (26.5)$$

نشان می‌دهد که مسیر جواب $y(t)$ کراندار است و بنابراین قضیه ۶.۳.۱ نتیجه می‌شود $y(t)$ ،
□ $t \in [t_0, \tau)$ وقتی $\tau \rightarrow +\infty$ وجود دارد.

قضیه ۴.۳.۵. فرض کنید y^* نقطه تعادل شبکه عصبی (۱۱.۵) و (۱۲.۵) باشد. جواب شبکه عصبی (۱۱.۵) و (۱۲.۵) با نقطه آغازین $y_0 \in S_0$ همگرای سراسری به جواب بهینه مسئله نیمه معین غیرخطی (۱۱.۵) و (۱۲.۵) است.

برهان. از (۲۶.۵) $y(t)$ کراندار است، بنابراین یک نقطه حدی $\hat{y} = (\hat{x}^T, \hat{z}^T, \hat{u}^T)^T$ و یک دنباله نزولی $\{t_k\}$ ($t_k \rightarrow +\infty$) وجود دارد به طوری که

$$\lim_{k \rightarrow \infty} y(t_k) = \hat{y}. \quad (۲۷.۵)$$

از (۲۶.۵)، وقتی که $\|y(t)\| \rightarrow \infty$ داریم $E(y) \rightarrow \infty$. پس مجموعه سطح

$$M(y_0) = \{y \in S_0 | E(y) \leq E(y_0)\}$$

کراندار است. همچنین $M(y_0)$ زیر مجموعه بسته \mathbb{R}^{n+m} و در نتیجه فشرده است. توجه داشته باشید که $\frac{dE(y(t))}{dt} \leq 0$ باعث می‌شود $y(t) \in M(y_0)$ برای هر $t \geq t_0$. با بکار بردن اصل تغییرناپذیری لسال در (۱۰.۳.۱) درمی‌یابیم که وقتی $t \rightarrow \infty$ ، $y(t) \rightarrow M$ که در آن بزرگترین مجموعه تغییرناپذیر در $\{y \in M(y_0) | \frac{dE(y(t))}{dt} = 0\}$ است. بنابراین از (۲۷.۵) داریم $\hat{y} \in M$ از (۲۶.۵) نتیجه می‌شود $\frac{dE(y(t))}{dt} = 0$ اگر و فقط اگر $\frac{dx}{dt} = 0$ و $\frac{dz}{dt} = 0$. پس \hat{y} یک نقطه تعادل شبکه عصبی (۱۱.۵) و (۱۲.۵) است. با استفاده از قضیه ۱.۳.۳، $\hat{y} = (\hat{x}^T, \hat{z}^T)^T$ نقطه بهینگی مسئله نیمه معین غیرخطی (۱.۵)–(۳.۵) است.

با جایگذاری $x^* = \hat{x}$ ، $z^* = \hat{z}$ و $u^* = \hat{u}$ در (۱۹.۵) تابع لیپانوف دیگری به شکل زیر تعریف می‌کنیم

$$\begin{aligned} \hat{E}(y) = & f(x) - f(\hat{x}) + \frac{1}{\alpha} [\|\bar{z}\|^2 - \|\hat{z}\|^2] + \frac{1}{\alpha} [\|\bar{u}\|^2 - \|\hat{u}\|^2] - (x - \hat{x})^T (\nabla f(\hat{x}) + \bar{A}\hat{z} + \nabla g(\hat{x})\hat{u}) \\ & - (z - \hat{z})^T \hat{z} - (u - \hat{u})^T \hat{u} + \frac{1}{\alpha} \|x - \hat{x}\|^2 + \frac{1}{\alpha} \|z - \hat{z}\|^2 + \frac{1}{\alpha} \|u - \hat{u}\|^2 \end{aligned}$$

$\hat{E}(y)$ پیوسته مشتق پذیر بوده و $\hat{E}(\hat{y}) = 0$. بنابراین، از (۲۷.۵)

$$\hat{E}_{k \rightarrow \infty}(x_k, z_k) = \hat{E}(\hat{x}, \hat{z}) = 0.$$

پس برای هر $\epsilon > 0$ یک $q > 0$ وجود دارد به طوری که برای تمام $t > t_q$ ، $\frac{d\hat{E}(y)}{dt} \leq 0$. مشابه فرایند بالا می‌توانیم اثبات کنیم که $\frac{d\hat{E}(y)}{dt} \leq 0$ از اینکه $t \geq t_q$ نتیجه می‌شود

$$\frac{1}{\alpha} (\|x(t) - \hat{x}\|^2 + \|z(t) - \hat{z}\|^2 + \|u(t) - \hat{u}\|^2) \leq \hat{E}(y) < \epsilon.$$

یعنی $\lim_{t \rightarrow \infty} x(t) = \hat{x}$ ، $\lim_{t \rightarrow \infty} z(t) = \hat{z}$ و $\lim_{t \rightarrow \infty} u(t) = \hat{u}$. بنابراین منحنی جواب شبکه عصبی (۱۱.۵) و (۱۲.۵) همگرای سراسری به یک نقطه KKT از مسئله نیمه معین غیرخطی (۱.۵)–(۳.۵) است. \square

نتیجه ۱.۳.۵. فرض کنید $D^* = \{(x^*, Z^*, u^*)\}$ ، مجموعه تمام نقاط KKT مسئله نیمه معین غیرخطی (۱.۵)–(۳.۵) باشد. آنگاه شبکه عصبی (۱۱.۵) و (۱۲.۵) پایدار مجانبی سراسری به نقطه تعادل $y^* = (x^{*T}, z^{*T}, u^{*T})^T$ است که $z^* = \text{vec}(Z^*)$.

۴.۵ مثال های عددی

در این بخش برای نشان دادن اعتبار مدل شبکه عصبی ارائه شده چندین مثال عددی آورده می شود. برای این مثال ها همچنین کارایی عددی شبکه عصبی (۱۱.۵) و (۱۲.۵) را با مقادیر مختلف ρ و نقاط آغازین مختلف y_0 بررسی می کنیم. شبیه سازی این مثال ها با استفاده از نرم افزار Matlab 7 و حل کننده معادلات دیفرانسیل معمولی ode45 انجام شده است.

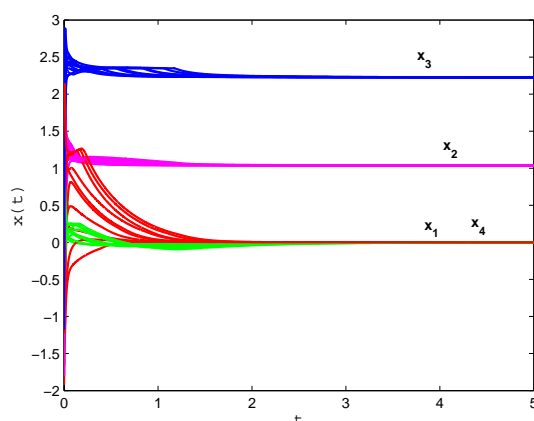
مثال ۱.۵ [۱۰۷] مساله زیر که از ترکیب مساله روسن-سوزوکی^{۱۱} با یک قید نیمه معین ساخته شده است را در نظر بگیرید

$$\begin{aligned} \text{minimize } f(x) &= x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4, \\ \text{subject to } g_1(x) &= x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 - 8 \leq 0, \\ g_2(x) &= x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 - 10 \leq 0, \\ g_3(x) &= 2x_1^2 + x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_2 - x_4 - 5 \leq 0, \end{aligned}$$

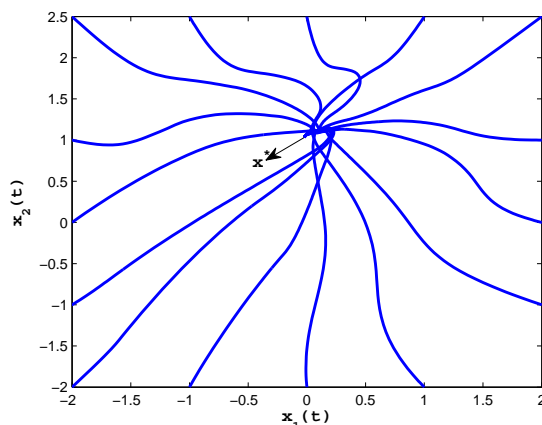
$$A(x) = \begin{pmatrix} x_2 + x_3 & 0 & 0 & 0 \\ 0 & 2x_4 & x_1 & 0 \\ 0 & x_1 & 2x_4 & 0 \\ 0 & 0 & 0 & x_2 + x_3 \end{pmatrix} \succeq 0.$$

جواب بهینه این مساله $(0, 1/0.380, 2/2272, 0)^T$ است. شکل ۱.۵ رفتار همگرایی منحنی های جواب $(x_1(t), x_2(t), x_3(t), x_4(t))^T$ را با 10^{-6} نقطه آغازین تصادفی نشان می دهد. شکل ۲.۵ و ۳.۵ به ترتیب نمودار فاز $(x_1(t), x_2(t))^T$ و $(x_3(t), x_4(t))^T$ را با 10^{-6} نقطه شروع مختلف نمایش می دهند. به سادگی می توان تصدیق کرد که نقطه آغازین چه درون ناحیه شدنی باشد یا بیرون آن، شبکه عصبی پیشنهاد شده در این فصل به جواب بهینه مساله همگراست. در این مساله همچنین تاثیر پارامتر ρ روی $\|x(t) - x^*\|^2$ را در شبکه عصبی (۱۱.۵) و (۱۲.۵) بررسی می کنیم. در شکل ۴.۵ مشاهده می کنیم که وقتی $\rho = 0.1$ است شبکه عصبی (۱۱.۵) و (۱۲.۵) کاهش نرم آهسته تری نسبت به زمانی که $\rho = 10^{-6}$ است تولید می کند. یعنی هر چه ρ بزرگتر باشد همگرایی $\|x(t) - x^*\|^2$ سریعتر است. شکل ۵.۵ رفتار همگرایی $\|x(t) - x^*\|^2$ مبتنی بر شبکه عصبی (۱۱.۵) و (۱۲.۵) را با 10^{-6} نقطه آغازین تصادفی نشان می دهد. سرانجام در این مثال با استفاده از لم (۳.۲.۲) شرط بهینگی (۲۶.۲) را بررسی می کنیم. مشاهده می شود که $\|Z^* A(x^*)\| = 0.2 \times 10^{-6}$.

^{۱۱}Rosen-Suzuki



شکل ۱.۵: رفتار همگرایی $(x_1(t), x_2(t), x_3(t), x_4(t))^T$ شبکه عصبی (۱۱.۵) و (۱۲.۵) با ۱۰ نقطه که به طور تصادفی در مثال ۱.۵ انتخاب شدند.

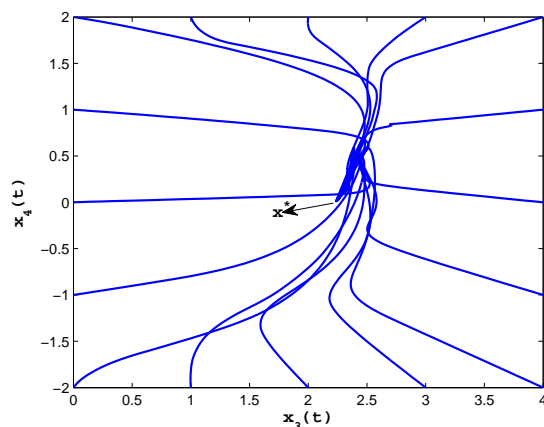


شکل ۲.۵: نمودار فاز $(x_1(t), x_2(t))^T$ شبکه عصبی (۱۱.۵) و (۱۲.۵) با ۱۵ نقطه آغازین در مثال ۱.۵.

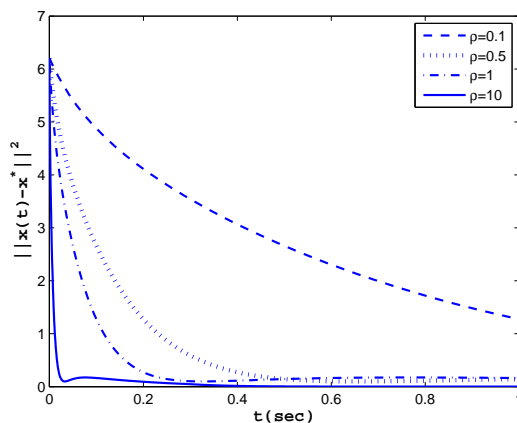
مثال ۲.۵. مساله نیمه معین غیرخطی زیر را در نظر بگیرید:

$$\begin{aligned} & \text{minimize } f(x) = x_1 + (x_2 - 1)^2 \\ & \text{subject to } A(x) = \begin{pmatrix} 1 & 0 & x_1 \\ 0 & 1 & x_2 \\ x_1 & x_2 & 1 \end{pmatrix} \succeq 0. \end{aligned}$$

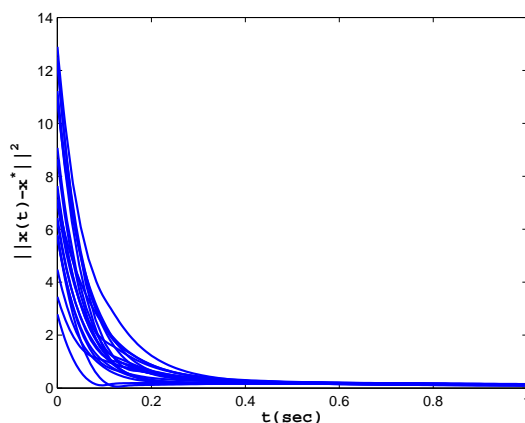
شبکه عصبی (۱۱.۵) و (۱۲.۵) را برای حل این مساله به کار می‌گیریم. نتایج شبیه‌سازی نشان می‌دهند که مسیرهای گذرا به جواب بهینه $x^* = (-0.7902, 0.6126)^T$ همگراست. شکل ۶.۵ نشان می‌دهد که منحنی مسیر شبکه عصبی همگرای سراسری است. در واقع قضیه ۳.۳.۵ تضمین می‌کند که مدل شبکه عصبی (۱۱.۵) و (۱۲.۵) همگرای سراسری به x^* است. شکل ۷.۵ همگرایی $\|x(t) - x^*\|^2$ مبتنی بر شبکه عصبی (۱۱.۵) و (۱۲.۵) را با ۲۰ نقطه آغازین



شکل ۳.۵: نمودار فاز $(x_3(t), x_4(t))^T$ شبکه عصبی (۱۱.۵) و (۱۲.۵) با ۱۵ نقطه آغازین تصادفی در مثال ۱.۵.

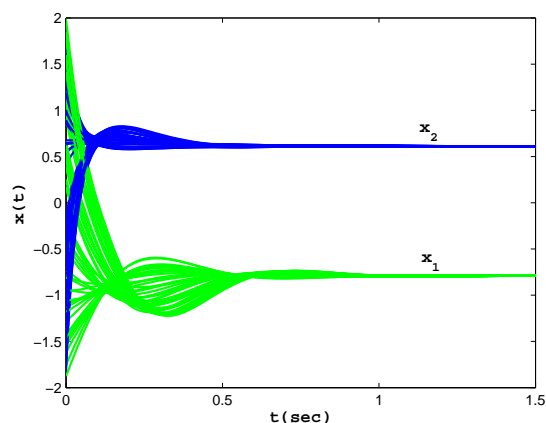


شکل ۴.۵: رفتار همگرایی $\|x(t) - x^*\|^2$ با $x_0 = (1, -1, 2, 1)^T$ در مثال ۱.۵.

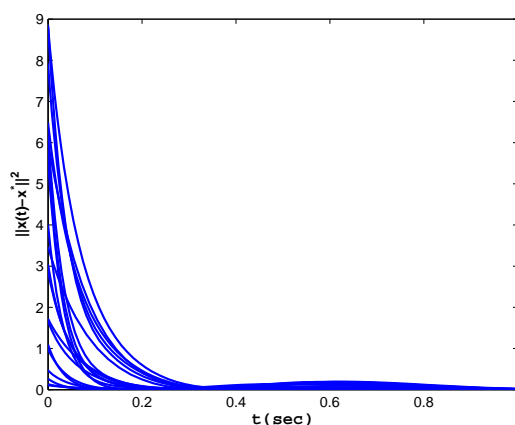


شکل ۵.۵: رفتار همگرایی $\|x(t) - x^*\|^2$ با ۲۰ نقطه آغازین تصادفی و $\rho = 1$ در مثال ۱.۵.

تصادفی نشان می دهد.



شکل ۶.۵: رفتار همگرایی $(x_1(t), x_2(t))^T$ مبتنی بر شبکه عصبی (۱۱.۵) و (۱۲.۵) با 5° نقطه آغازین تصادفی در مثال ۲.۵.



شکل ۷.۵: رفتار همگرایی $\|x(t) - x^*\|^2$ مبتنی بر شبکه عصبی (۱۱.۵) و (۱۲.۵) با 2° نقطه آغازین تصادفی و $\rho = 5$ در مثال ۲.۵.

۵.۵ نتایج و پیشنهادات

مسائل برنامه‌ریزی نیمه معین در دهه‌های اخیر بسیار مورد توجه قرار گرفته‌اند و کاربردهای بسیار گسترده‌ای یافته‌اند. ما در این رساله یک مدل شبکه عصبی برای حل مساله نیمه معین خطی بر اساس شرایط بهینگی KKT ارائه دادیم. شبکه عصبی پیشنهاد شده ساختاری ساده دارد. علاوه بر آن نشان دادیم که شبکه عصبی ارائه شده پایدار لیاپانوف و همگرای سراسری به جواب بهینه مساله نیمه معین خطی است. همچنین کارایی مدل شبکه عصبی ارائه شده با حل مثال‌های متنوع نشان داده شد. همچنین یک مدل شبکه عصبی گرادیانی مبتنی بر تابع فیشر- برمیچر برای حل مساله برنامه‌ریزی نیمه معین خطی ارائه دادیم. نشان دادیم که مدل شبکه عصبی ارائه شده پایدار مجانبی و همگرای سراسری به جواب بهینه مساله نیمه معین اصلی است. با ارائه مثال‌هایی کارایی مدل شبکه عصبی پیشنهاد شده تایید شد. از آنجایی که مساله برنامه‌ریزی نیمه معین کاربردهای گسترده‌ای دارد، مدل شبکه عصبی ارائه شده و نتایج بدست آمده در این رساله هم در نظریه و هم کاربرد بسیار مورد توجه است. با توجه به این که تاکنون مدل شبکه عصبی برای حل مساله نیمه معین غیرخطی ارائه نشده است، در این رساله برای اولین بار یک مدل شبکه عصبی مبتنی بر شرایط KKT برای حل این مسئله ارائه شده است. این مدل شبکه عصبی دارای ساختاری ساده است. همچنین شبکه عصبی ارائه شده پایدار به مفهوم لیاپانوف و همگرای سراسری به جواب مساله نیمه معین غیرخطی است. با توجه به کاربردهای بسیار گسترده مسئله نیمه معین غیرخطی نتایج بدست آمده در این رساله می‌تواند مورد توجه قرار بگیرد.

به عنوان پیشنهادی برای کارهای آتی می‌توان به ارائه مدل‌های شبکه عصبی بر اساس فاصله دوگانی و شبکه عصبی گرادیانی بر مبنای توابع شایسته مختلف برای مساله نیمه معین خطی و مساله نیمه معین غیرخطی اشاره کرد.

مراجع

- [۱] ضمیری، آذر. شبکه‌های عصبی مصنوعی، پایان نامه کارشناسی ارشد، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد، آبان ۱۳۸۷.
- [۲] قماش‌لنگرودی، عباس، حل مسائل بهینه‌سازی با استفاده از شبکه‌های عصبی تصویر، پایان نامه کارشناسی ارشد، دانشکده علوم پایه، دانشگاه تربیت معلم سبزوار، آبان ۱۳۸۳
- [3] F. Alizadeh, Interior point methods in semidefinite programming with applications to combinatorial optimization, *SIAM J. Optimization*, 5(1995) 13-51.
- [4] F. Alizadeh, J.P. A. Haeberly and M. L. Overton, Complementarity and nondegeneracy in semidefinite programming, *Mathematical Programming*, 77 (1997) 111-128.
- [5] F. Alizadeh, Optimization over the positive-definite cone: interior point methods and combinatorial applications, in *Advances in Optimization and Parallel Computing*, E Pardalos, ed., North-Holland, the Nether lands, (1992).
- [6] E. Anderson and P. Nash, *Linear Programming in Infinite Dimensional Spaces*, John Wiley and Sons, New York, (1987).
- [7] M.M. Arefi, J. Zarei and H.R. Karimi, Adaptive output feedback neural network control of uncertain non-affine systems with unknown control direction, *Journal of the Franklin Institute* 351 (2014) 4302–4316
- [8] M. S. Bazaraa, J.J. Jarvis and H.D. Sherali, *Linear Programming and network flows*, 4th Edition, Applied Mathematics, January (2010).
- [9] M. S. Bazaraa, H. D. Sherali and C.M. Shetty, *Nonlinear Programming*. John Wiley Sons, Inc., 3rd ed., (2005).
- [10] R. Bellman and K. Fan, On systems of linear inequalities in Hermitian matrix variables, in *Convexity*, V. L. Klee, ed, Symposia in Pure Mathematics, American Mathematical Society Providence, RI, 7(1963)1-11.

-
- [11] A. Ben-Tal and M.P. Bendse, A new method for optimal truss topology design, *SIAM Journal on Optimization*, 3(1993)322-358.
- [12] J. M. Borwein and H. Wolkowicz, Characterization of optimality for the abstract convex program with finite dimensional range, *Journal of the Australian Mathematical Society* 30 (1981)390- 411.
- [13] A. Bouzerdoum and T.R. Pattison, Neural network for quadratic optimization with bound constraints. *IEEE Transactions on Neural Networks* 4 (1993)293–304.
- [14] S. Boyd, L. E Ghaoui, E. Feron, V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, SIAM, Philadelphia(1994).
- [15] J.-S. Chen, C.-H. Ko, S.-H. Pan, A neural network based on the generalized Fischer-Burmeister function for nonlinear complementarity problems, *Information Sciences* 180 (2010) 697-711.
- [16] B. Craven and B. Mond, Linear programming with matrix variables, *Linear Algebra and its Applications*, 38 (1981)73-80.
- [17] R. Correa, C. Hector Ramirez, A global algorithm for solving nonlinear semidefinite programming. *SIAM Journal on Control and Optimization*, 15(2004)303-318.
- [18] J.Cullum, W. Donath and P. Wolef, The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices, *Mathematical Programming Studies*, 3(1975)35-55.
- [19] R. J. Duffin, Infinite programs, in *Linear Inequalities and Related Systems*, H. W. Kuhn and A. W. Tucker, eds., Princeton University Press, Princeton, NJ(1956)157-170.
- [20] S. Effati and M. Baymani, A new nonlinear neural network for solving convex nonlinear programming problems. *Applied Mathematics and Computation* 168(2)(2005)1370 – 1379.
- [21] S.Effati, A. Ghomashi and M. Abbasi, A novel recurrent neural network for solving mlcps and its application to linear and quadratic programming. *Asia-Pacific Journal of Operational Research*, 28(4)(2011)523–541.
- [22] S. Effati, A. Ghomashi and A.R. Nazemi, Application of projection neural network in solving convex programming problems. *Applied Mathematics and Computation*, 188(2007)1103 – 1114.

- [23] S. Effati, A.R. Nazemi, Neural network and its application for solving linear and quadratic programming problems, *Applied Mathematics and Computation* 172 (2006) 305-331.
- [24] M. K. H. Fan, A quadratically convergent algorithm on minimizing the largest eigenvalue of a symmetric matrix, *Linear Algebra and its Applications*, 188(1993)231-253.
- [25] M. K. H.Fan and B. Nekooie, On Minimizing the Largest Eigenvalue of a Symmetric Matrix, *Linear Algebra and its Applications*, 214(1995)225-246.
- [26] B. Fares, P. Apkarian, D. Noll, An augmented Lagrangian method for a class of LMI-constrained problems in robust control theory, *International Journal of Control*, 74(2001)348-360.
- [27] B. Fares, D. Noll, P. Apkarian, Robust control via sequential semidefinite programming, *SIAM Journal of Control and Optimization*, 40(2002)1791-1820.
- [28] A. Fischer, A special Newton-type optimization methods, *Optimization* 24 (1992) 269–284.
- [29] R. Fletcher, Semidefinite matrix constraints in optimization, *SIAM Journal on Control and Optimization*,23(1985) 493-513.
- [30] R. M. Freund, *Introduction to Semidefinite Programming (SDP)*, Massachusetts Institute of Technology, March 2004.
- [31] F. Giannessi and A. Maugeri, *Variational Analysis and Applications. Nonconvex Optimization and Its Applications*. Springer, 2005.
- [32] Q. Han, L.-Z. Liao, H. Qi and L. Qi, Stability analysis of gradient-based neural networks for optimization problems, *Journal of Global Optimization* 19 (2001) 363-381.
- [33] D.O.Hebb, *The Organization of Behavior: A Neuropsychological Theory*. John Wiley and Sons, Inc., New York, 1949.
- [34] C. Helmberg, *Semidefinite programming for combinatorial optimization*, Konrad-Zuse-Zentrum fur Informationstechnik Berlin , (2000).
- [35] C.Helmberg, F. Rendl and H.Wolkowicz, An interior point method for semidefinite programming. *SIAM Journal of Optimization*, 6(1996)342-361.
- [36] J.J.Hopfield, Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79(8)(1982)2554–2558.

- [37] J.J. Hopfield and D.W. Tank, Neural computation of decision in optimization problems, *Biological Cybernetics* 52 (1985) 141-152 .
- [38] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, U.K., 1991.
- [39] X. Hu, J. Wang, A recurrent neural network for solving nonlinear convex programs subject to linear constraints, *IEEE Transactions on Neural network* 16(2005) 379-386.
- [40] X. Hu, J. Wang, A recurrent neural network for solving a class of general variational inequalities, *IEEE Transactions on Systems, Man, and Cybernetic s-B* 37 (2007) 528-539.
- [41] L. Hurwicz, Programming in linear spaces, in *Studies in Linear and Nonlinear Programming*, II, K. J. Arrow, L. Hurwicz, and H. Uzawa, eds., Stanford University Press, Stanford, CA, (1958) 4-102.
- [42] D. Jiang, J. Wang, A recurrent neural network for real-time semidefinite programming, *IEEE Transactions on neural networks*, 10(1999)81-93.
- [43] F. Jarre, An interior-point method for minimizing the maximum eigenvalue of a linear combination of matrices, *SIAM Journal of Control and Optimization*, (1993)1360-1377.
- [44] N. Kalouptisidis, *Signal Processing Systems, Theory and Design*, Wiley, New York, (1997).
- [45] A. Kamath and N. Karmakar, Continuous approach to compute upper bounds in quadratic maximization problems with integer constraints, in *Recent Advances in Global Optimization*, C. Floudas and P. Pardalos, eds., Princeton University Press, Oxford, UK, (1992) 125-140.
- [46] C. Kanzow and C. Nagel, Semidefinite programs: new search directions, smoothing-type methodes, and numerical results, *SIAM Journal of Optimization*(2002)1-23.
- [47] N. Karmakar, A new polynomial-time algorithm for linear programming, *Combinatorica*, 90(1984)373- 395.
- [48] H.R. Karimi and H. Gao, New delay-dependent exponential synchronization for uncertain neural networks with mixed time delays, *IEEE Transactions on Systems, Man and Cybernetics* 40 (2010) 173–185.
- [49] M.P. Kennedy and L.O. Chua, Neural network for nonlinear programming, *IEEE Transactions on Circuits and Systems* 35(1988) 554-562.

- [50] H.K. Khalil, *Nonlinear Systems*. Prentice Hall PTR, 3rd ed. , 2002.
- [51] B. Kheirfam, Primal- dual interior-point algorithm for semidefinite optimization based on a new kernel function with trigonometric barrier term, *Springer Science+Business Media, LLC* 61(2012)659-680.
- [52] C.-H. Ko, J.-S. Chen and C.-Y. Yang, Recurrent neural networks for solving second-order cone programs, *Neurocomputing* 74 (2011) 3646-3653.
- [53] M. Kojima, S. Kojima and S. Hara, *Linear algebra for semidefinite programming*, Tech. report, Dept. Mathematical and Computing Sciences, Tokyo Institute of Technology, Oh-Okayama, Meguro-ku, Tokyo 152, Japan, Oct. (1994).
- [54] M. Kojima, M. shida and S. Hara, Interior point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. *SIAM Journal of Optimization*. 7(1997)86-125.
- [55] J.LaSalle and S. Lefschetz, *Stability by Lyapunov's Direct Method with Applications*, Academic Press, New York, (1961).
- [56] Y. Leung, X.B. Gao and K.Z. Chen, A dual neural network for solving entropy-maximising models, *Environment and Planning*, 36(2004)897-919.
- [57] L.-Z. Liao, H. Qi and L. Qi, Solving nonlinear complementarity problems with neural networks: a reformulation method approach, *Journal of Computational and Applied Mathematics* 131 (2001) 342-359.
- [58] D.R. Liu, D.Wang and X. Yang, An iterative adaptive dynamic programming algorithm for optimal control of unknown discrete-time nonlinear systems with constrained inputs, *Information Sciences* 220 (2013) 331-342.
- [59] D. Luenberger and Y.Ye, *Linear and Nonlinear Programming*. Springer, 3rd ed. , (2008).
- [60] C.Y.Maa and M.A.Schanblatt, A two-phase optimization neural network. *IEEE Transactions Neural Networks* 3(1992)1003–1009.
- [61] O.L. Mangasarian, Equivalence of the complementarity problem to a system of nonlinear equations, *SIAM Journal of Applied Mathematics* 31(1976)89-92.
- [62] M. Warren and W.Pitts, A logical calculus of the ideas immanent in nervous activit, *Bulletin of Mathematical Biophysics* 5(1943)115–133.

- [63] R.K. Miller and A.N. Michel, Ordinary differential equations, Academic Press, New York, (1982).
- [64] M.L.Minsky and S.Papert Perceptrons: An Introduction to Computational Geometry. Mit Press, 1972.
- [65] M.Moghaddas and S.Effati, A novel recurrent neural network based on ncp function for solving convex nonlinear optimization problems. in The 5th Iranian Conference on Applied Mathematics, Hamadan, Iran, September 2013. Bu-Ali Sina University.
- [66] M.Moghaddas and M.Eshaghnejad, A recurrent neural network for solving nonconvex nonlinear optimization problem. in The 44th Annual Iranian Mathematics Conference, Mashhad, Iran, August 2013. Ferdowsi University of Mashhad.
- [67] B. Mohar and S.Poljak, Eigenvalues in combinatorial optimization, in Combinatorial and Graph-Theoretical problems in Linear Algebra, R.A.Brualdi, S.Friedland, and V.Klee, eds., Springer-Verlag, New York, Berlin, 50(1993)107-151.
- [68] A. R. Nazemi, A dynamical model for solving degenerate quadratic minimax problems with constraints, Journal of Computational and Applied Mathematics, 236 (2011) 1282-1295.
- [69] A.R. Nazemi, Solving general convex nonlinear optimization problems by efficient neurodynamic model, Engineering Applications of Artificial Intelligence, 26(2013)685-696.
- [70] A.R. Nazemi and M.Nazemi, A Gradient-Based Neural Network Method for Solving Strictly Convex Quadratic Programming Problems, Cognitive Computation 6 (2014) 484–495.
- [71] Y. Nesterov and A. Nemirovsky, A general approach to polynomial-time algorithms design for convex programming, Technical report Center Economy and Mathematic Institute, USSR Acad. Sci., Moscow, USSR, (1988).
- [72] Y. Nesterov and A.Nemirovsky, Interior-point polynomial methods in convex programming, Studies in Applied and Numerical Mathematics, SIAM, Philadelphia, 13(1994).
- [73] T.M.Ortega and W.C. Rheinboldt, Iterative solution of nonlinear equations in several variables, Academic Press, New York, (1970).
- [74] M.Overton, On minimizing the maximum eigenvalue of a symmetric matrix, SIAM Journal of Matrix Analysis and Applications, 9(1988) 256-268.

- [75] G. Patakoï, On the multiplicity of optimal eigenvalues, Tech. report MSRR-604, Graduate School of Industrial Administration, Camegie Mellon University, Pittsburgh, PA, (1994).
- [76] Perko, L. Differential Equations and Dynamical Systems. Springer, 3rd ed. (2001).
- [77] F. Pukelsheim, Optimal Design of Experiments, Wiley, (1993).
- [78] A. Quarteroni, R. Sacco and F. Saleri, Numerical mathematics, volume 37 of Texts in Applied Mathematics, Springer-Verlag, Berlin, 2nd ed., (2007).
- [79] M. Ramana, An exact duality theory for semidefinite programming and its complexity implications, *Mathematical Programming*, 77(1997)129-162.
- [80] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, second ed., 1970.
- [81] A. Rodriguez-Vazquez, R. Dominguez-Castro, A. Rueda, J.L. Huertas and E. Sanchez-Sinencio, Nonlinear switched capacitor neural networks for optimization problems. *IEEE Transactions on Circuits and Systems*, 37(1990)384–398.
- [82] F. Rosenblatt, *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*. Report (Cornell Aeronautical Laboratory). Spartan Books, 1962.
- [83] F. Rosenblatt, a probabilistic model for information storage and organization in the brain, *Neurocomputing foundations of research*, 66(1988)89–114.
- [84] D.E. Rumelhart, G.E. Hinton and R.J. Williams, Learning representations by back-propagating errors. *Nature*, 323(1986)533–536.
- [85] J.H. Sun, J.S. Chen and C.H. Ko, Neural networks for solving second-order cone constrained variational inequality problem, *Computational Optimization and Applications*, 51 (2012) 623-648.
- [86] D.W. Tank and J.J. Hopfield, Simple neural optimization network: an A/D converter, signal decision circuit, and a linear programming circuit, *IEEE Transactions on Circuits and Systems* 33 (1986) 533-541.
- [87] P. Tseng, Merit functions for semi-definite complementarity problems, *Mathematical Programming*, 83(1998)159-185.
- [88] L. Vandenberghe and S. Boyd, Semidefinite programming, *SIAM Review*, (1996)45-49.

-
- [89] L. Vandenberghe and S. Boyd, Primal-dual potential reduction method for problems involving matrix inequalities, *Mathematical Programming, Series B*, 69(1995)205-236.
- [90] G.Q. Wang and Y.Q. Bai, Primal-dual interior-point algorithms for semidefinite optimization based on a simple kernel function *Journal of Mathematical Modeling and Algorithms*, 4(2005)409-433.
- [91] Z. Wen, D. Goldfarb and W. Yin, Alternating direction augmented Lagrangian methods for semidefinite programming, *Mathematical Programming and Computation*, 2(2010)203-230.
- [92] H. Wolkowicz, Some applications of optimization in matrix theory, *Linear Algebra and Applications* (1981)101-118.
- [93] P. Werbos, *Beyond Regression, New Tools for Prediction and Analysis in the Behavioral Sciences*. Ph.D. thesis, Harvard University, Cambridge, MA, 1974.
- [94] B. Widrow, Generalization and information storage in networks of adaline neurons. in *Self Organizing Systems*, Spartan Books(1959)435-461.
- [95] B. Widrow and M.E. Hoff, Adapting switching circuits. in 1960 ire wescon Convention Record, no. 4, New York, NY (1960)96-104.
- [96] B. Widrow and M.A. Lehr, 30 years of adaptive neural networks: perceptron, madaline, and backpropagation. *Proceedings of the IEEE*, 78(1990)1415-1442.
- [97] B. Widrow and R. Winter, Neural nets for adaptive filtering and adaptive pattern recognition. *Computer*, 21(1988)25-39.
- [98] H.X. Wu, H.Z. Luo, X.O. Ding and G. Chen, Global convergence of modified augmented Lagrangian methods for nonlinear semidefinite programming, *Computational Optimization and Applications* 56(2013) 531-558.
- [99] H. X. Wu, H. Z. Luo and J. F. Yang, Nonlinear separation approach for the augmented Lagrangian in nonlinear semidefinite programming, *Journal of Global Optimization* 59(2014)695-727.
- [100] A.L. Wu, S.P. Wen and Z.G. Zeng, Synchronization control of a class of memristor-based recurrent neural networks, *Information Sciences* 183(2012) 106-116.

- [101] Y.Xia, G.Feng and J.Wang, A novel recurrent neural network for solving nonlinear optimization problems with inequality constraints. IEEE Transactions. Neural Networks, 19(2008)1340–1353.
- [102] Y. Xia, H. Leung and J. Wang, A projection neural network and its application to constrained optimization problems, IEEE Transactions on Circuits and Systems-I 49 (2002) 447-458.
- [103] Y. Xia, H. Leung and J. Wang, A general projection neural network for solving monotone variational inequalities and related optimization problems, IEEE Transactions on Neural Networks 15(2004) 318-328.
- [104] Y. Xia and J. Wang, A recurrent neural network for solving nonlinear convex programs subject to linear constraints, IEEE Transactions on Neural Networks 16(2005) 379-386.
- [105] Y.Xia and J.Wang, A general methodology for designing globally convergent optimization neural networks. IEEE Transactions on Neural Networks 9(1998)1331–1343.
- [106] Y.Xia and J.Wang, A recurrent neural network for nonlinear convex optimization subject to nonlinear inequality constraints. IEEE Transactions on Circuits Systems. I, Reg. Papers, 51(2004)1385–1394.
- [107] H. Yamashita and H. Yabe, A primal-dual interior point method for nonlinear semidefinite programming. Institute Statistical Mathematics. Cooper. Res. Report 203(2007)296-317.
- [108] Q.Yang and H.Yu, Discretization method for semi-definite programming, Computers and Mathematics with Applications, 48 (2004) 1937-1945.
- [109] Y.Yang and J.Cao, The optimization technique for solving a class of non-differentiable programming based on neural network method, Nonlinear Analysis: Real World Applications 11 (2010) 1108-1114
- [110] M. Yashtini, A. Malek, Solving complementarity and variational inequalities problems using neural networks, Applied Mathematics and Computation 190 (2007) 216-230.
- [111] T. Yoshikawa, Foundations of Robotics: Analysis and Control, MIT Press, Cambridge, MA, 1990.
- [112] J.Zabczyk, Mathematical control theory: an introduction. Systems control. Birkhäuser, 1992.

- [113] G.D. Zhang, Y. Shen, Q. Yin and J.W. Sun, Global exponential periodicity and stability of a class of memristor-based recurrent neural networks with multiple delays, *Information Sciences* 232 (2013) 386-396.

Aabstract

Semidefinite problems has been very much considered in recent decades and includes a very large area of convex optimization problems. Many of the important issues in system theory and control, structural optimization, statistics, and other fields can be formulated as a Semidefinite problem. So far, there are several ways to solve a Semidefinite problems. However, the use of neural networks to solve this problem has not been reported except in a small number.

In this thesis, two neural network models will be presented to solve a semidefinite programming problem. We will show that the proposed equilibrium point of the neural network and the optimal solution of the semidefinite problem are equivalent. The stability and convergence of the proposed neural network models will also be proved. Also, for the first time, we present a neural network model for solving a nonlinear semidefinite problem. The stability and global convergence prove to be the optimal solution of the main problem. By providing examples we confirm the theoretical results.

Keywords:Neural network; Semidefinite programming; Convex programming; Globally convergent.



Shahrood University of Technology

Faculty of Mathematical Sciences

PhD Thesis in: Control and Optimization

**Solving a class of semidefinite programming
problems using neural networks**

By: Asiye Nikseresht

Supervisor

Dr. Alireza Nazemi

January 2019