```
Lecture-by-lecture list of topics
EECS 598-006 W20 "Optimization methods for ..."

1 1/09
        Ch. 0 Course policies

        Ch. 1 Applications (1)
                1.0 introduction (read)
                1.1 linear programming
                        compressed sensing with l1 norm
                        minmax sparse filter design (read)
                        MRI RF pulse design (read)
                convex relaxation
                        sublevel sets, quasiconvex convex functions
                        convex envelope
                1.2 quadratic problems
                        LS, regularized LS, finite differences
                        constrained LS (read)
                                analytical solution for diagonal case

                1.3 strictly convex smooth problems
                        edge-preserving regularization
                        m-estimation for robust regression

2 1/14
                                robust regression example

                1.4 convex composite problems
                        l1 regularization
                        LASSO - feature selection
                        sparse approximation
                        signal models: synthesis and analysis
                                wave+spike example
                        denoising using sparse synthesis
                                unitary case
                        compressed sensing - synthesis regularizer
                        LASSO via GP using x = u-v
                        elastic net regularizer (Read)
                        analysis sparsity / total variation (TV)
                        1D TV as a LASSO problem (Read)
                        corner rounding (Read)
                        proximal operators - complex case (Read)

                1.5 non-smooth problems
                        robust regression using l1
                        binary classification (Read)
                        union of subspaces - unsupervised

3 1/16

        Ch. 2 Applications (2)
                2.1 signal processing applications
                        patch-based regularization
                                synthesis form
                                analysis form
                                aggregate (global) vs local sparsity
                        dictionary learning
                        transform learning
                                regularized version (Read)
                        filter learning
                        blind deconvolution
                        phase retrieval (Read)


4 1/21

                2.2 machine-learning applications
```

```
                        low-rank approximation / matrix factorization
                        low-rank matrix completion
                        matrix sensing / recovery

        Ch3: Gradient-based optimization
                3.1 Lipschitz continuity
                        properties

5 1/23

                        bounds / majorization
                        relate to Hessian
                        edge-preserving regularizer: Lipschitz constant
                3.2 gradient descent (GD)
                        step size
                        convergence rates

6 1/28 [claire lin, due to sedona workshop]

                3.3 preconditioned steepest descent (PSD)
                        preconditioning
                        descent direction
                                complex case for LS (Read)
                3.4 descent direction for edge-preserving regularizer: complex case
                        descent direction proof (Read)
                        Lipschitz constant conjecture (Read)
                        practical Lipschitz constant for edge-preserving regularizer (Cover!)
                        PGD step size
                        SD vs CG and inner products

                3.5 General inverse problems cost function
                        efficient line search (Read?)

7 1/30 [steven whitaker, due to ipam workshop]

                3.6 convergence rates of PGD, PSD, PCG
                        heavy ball method
                                heavy ball convergence analysis (Read)
                        S-Lipschitz continuity
                        PGD convergence theorem for S-Lipschitz convex functions
                        Lipschitz constant units
                        Nesterov's FGM with (preconditioned) gradient step

                3.7 First-order methods: general and fixed step
                        FGM is FO, FGM rates

8 2/04

                        OGM, OGM', bounds and optimality
                        OGM worst-case functions, bound tightness

                3.8 Logistic classifier design
                        cost function is like an inverse problem
                        Lipschitz constant via properties
                        example of GD / FGM / OGM
                        Adaptive restart of FGM / OGM
                                Strongly convex functions

reading: 11.8 (CG) 11.9 (QN) 11.10 (BCD) by 2/13 (should have started earlier)

9 2/06

                        nonlinearity in machine learning

                3.10 summary
                        history of first-order methods
                        preview of GFOM / OGM line search

                3.9 1D finite differences: demo/diff1, @view etc.
```

```
10 2/11

                        2D finite differences: in-class task
                                adjoint tests for LinearMaps
                                julia call-by-reference ?

11 2/13

        Ch. 4: Majorize-minimize (MM) methods
                4.0 Intro / Application examples
                4.1 Majorization principle / sandwich inequality
                        Algebraic properties
                        Quadratic majorizer when gradient is S-Lipschitz smooth
                        connection to PGD
                4.2 Applications
                        low-rank matrix completion via MM
                        LASSO / sparse regression / compressed sensing (l2+l1)

12 2/18

                        ISTA / PGM
                        diagonal majorizer
                Convexity majorizers (separability)
                        general case
                        LASSO example
                        exponential loss example (read)

reading (optional - not in W20 since ch4 now more complete):
        14.1 intro, skip 14.1.4
        14.5 surrogate design
        14.6.6 monotone line searches

13 2/20

                Poisson data (MLEM)
                Line search with Huber's majorizer
                        1D MM approach

14 2/25
                        rationale in terms of 1D mean/median/robust fair
                        Huber's conditions and uniqueness (read)
        [Exam review based on p11,14,24,33,34 of w20exam1 student problems]

** 2/26 Exam1 Wed. 6-8 PM

15 2/27
                Huber hinge loss
                        Lipschitz constant
                        optimal quadratic majorizer for s >= -1

                4.3 Acceleration methods (over-relaxation)
                        LASSO example

* 3/3 3/5 break

16 3/10

                4.4 Summary and LASSO recap

        Ch. 5: Proximal methods
                5.1 proximal operator definition
                        example: SVHD
                        example: hard thresholding, pocs

17 3/12 - Canceled due to corona virus

18 3/17
```

```
                proximal point algorithm, cf MM
                5.2 proximal gradient method (PGM)
                        cf MM
                        PGM convergence rate O(1/k)
                                linear rate (read)
                                strongly convex f
                        PGM with line search (read)

19 3/19

                PGM revisit

                5.3 accelerated proximal methods
                        FPGM (fast proximal gradient method) = FISTA
                        POGM (proximal optimal gradient method)
                        inexact compute of proximal operators
                        related methods

                5.4 applications
                        binary classifier with l1 regularizer

20 3/24

                        MRI compressed sensing with ODWT

        Ch. 6: Alternating minimization methods (BCD, BCM)
        6.1 SP applications
                CS using synthesis sparsity:
                        two-block BCM
                        two-block BCD
                CS using analysis sparsity: two-block BCD/BCM

21 3/26

                Sparse coding using multi-block BCM (aka CD)
                        relate to MP/OMP (brief in w19, not in w10)
                        CD code
                CD update of x for CS with synthesis sparsity
                        (group task in w19, not in w20)
                Sparse coding for tight frame via 2-block BCM
                        (group notebook in w19, not in w20)
                Wavelet transform overview (in w19, not in w20)
                Patch-based regularization using sparsifying transform

22 3/31

                Sparsifying transform learning via two-block BCM
                        square case via orth. procrustes
                        non-square case via MM (Read in w20)
                        non-square case by weighted 0-norm (Read in w20)
                        example of 1D filter learning
                        memory efficient implementation (Read in w20)

                Dictionary learning
                        two-block update of D,Z (brief)
                        multi-block update of d_k,c_k ala SOUP
                        in-class task using SOUP for wave+spike (not w20)
                        joint update of dk and ck (Read)


23 4/02
                6.2 ML applications
                        Low-rank matrix approximation for large problems via BCM
                        Non-negative matrix factorization & sparsity
                        fused LASSO (brief/read)
                        alt min for 0-norm in biconvex form (brief)

                6.3 convergence properties
```

```
                6.4 more about BCM (didn't do in W20)
                        relating to GD
                        VARPRO
                        1D TV failure to converge

24 4/07

        Ch. 7: Duality / Lagrangian / ADMM
                7.0 variable splitting overview
                7.1 convex conjugate
                        properties & examples
                7.2 Method of Lagrange multipliers
                        Lagrange dual
                        Properties and relation to convex conjugate
                        weak and strong duality
                        Using dual problem to solve the primal problem
                7.3 Augmented Lagrangian method
                        ADMM


25 4/09

                        binary classifier with hinge via ADMM
                        ADMM in general / convergence
                        linearized AL method (LALM)
                        augmented ADMM
                        primal-dual hybrid gradient (PDHG) / Chambolle-Pock
                        "near circulant splitting method"

26 4/14

        Ch. 8 Stochastic gradient / subgradient methods
        8.1 Subgradient method
                Subdifferential / subgradient and properties
                Convergence for diminishing step sizes
                Normalized subgradient method
                Convergence for constant step size factors
                Projected subgradient method

                Naveen Murthy: SGM
        8.2 Example: hinge loss with 1-norm regularizer

27 4/16

        8.3 Incremental (sub)gradient method
        8.4 Stochastic gradient method
                minibatches
                convergence analysis
                variance reduction: SVRG, SAGA, ...
                momentum
                adaptive step sizes
                restart
                Example: ordinary linear LS
        8.5 Example: X-ray CT reconstruction

28 4/21

        Ch. 9: Misc. topics
        Review based on binary classifier design
                partly inspired by student exam2 questions

---- below here from W19 -------

        PGM as alternating between GD and denoising
        Patch transform sparsity as related to variational CNN methods
        Overview of deep learning for medical imaging
```